OLIGOMERIZATION OF THE LysR-TYPE TRANSCRIPTIONAL REGULATORS IN *Escherichia coli*

A Dissertation

by

GWENDOWLYN SUE KNAPP

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2008

Major Subject: Biochemistry

OLIGOMERIZATION OF THE LysR-TYPE TRANSCRIPTIONAL REGULATORS IN *Escherichia coli*

A Dissertation

by

GWENDOWLYN SUE KNAPP

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	James C. Hu
Committee Members,	Gregory D. Reinhart
	Ryland Young
	Michael Benedik
Head of Department,	Gregory D. Reinhart

May 2008

Major Subject: Biochemistry

ABSTRACT

Oligomerization of the LysR-Type Transcriptional Regulators

in *Escherichia coli.* (May 2008)

Gwendowlyn Sue Knapp, B.S., Purdue University

Chair of Advisory Committee: Dr. James C. Hu

Protein-protein interactions regulate and drive biological processes and understanding the assembly of these interactions is important. The LysR-Type Transcriptional Regulators (LTTRs) are a large family of transcriptional regulators found in prokaryotes. I have used the LTTRs as a model for protein specificity. In order to understand a residue's contribution to oligomerization, alanine-scanning mutagenesis was used to probe the contribution of residues identified from in silico analysis of two proteins: OxyR and CynR. The contribution of the residues to oligomerization was characterized using λcI repressor fusions. In OxyR, seven residues were identified as hot spots. Moreover, these hot spots are not especially conserved. The interaction surface of OxyR was mapped onto a multiple sequence alignment of the LTTR family. This mapping identified putative contacts in the CynR regulatory domain dimer interface. Combined with the in vivo testing, three residues were identified as hot spots. The residues identified in OxyR and CynR do not overlap. To investigate the assembly of the LTTRs I used a negative-dominance assay with λcI repressor fusions. Taken together, I show that the LTTRs in *E. coli* K-12 are mostly specific in their interactions.

DEDICATION

To my parents, Marlene and Wally Knapp, because they always believed in me.

ACKNOWLEDGEMENTS

I spent a lot of time by myself writing this dissertation, but the work was done with many people. I am greatly indebted to my boss, Dr. Jim Hu. The amount of time it took for me to learn to "Wait. Back up, slow down and explain the question" was long and the patience and belief he had that I could do "this" was enormous. The lesson of doing whatever it takes to answer the question (a good question, mind you) is priceless and the flexibility and enthusiasm for me is appreciated. To Jim, I say thank you.

I made many friends during my time at TAMU. Without my classmates, Tony Reeves, Patrick Frantom, Matt Bramlett and Jason Quinlan, my first year would have been horrible. I'm honored to call them classmates and now, colleagues. To Patrick, I say "What about this??" I thank my close friends Kari Halbig, Carrie Langlais and Adrienne Zweifel for their listening ears about all topics, relevant or not. I would also like to acknowledge Matt Champion, for being a partner in crime and a friend. Flaming wheel of death anyone?

I must also recognize Dr. Jerry Tsai, my collaborator. His patience in helping me code was unending and my scripts would not have come to light without him. He was also a great friend with great advice. I am thankful that I had his help and guidance. I also must thank Dr. Debby Siegele. Debby is a fantastic teacher and I am indebted to her for her patience and encyclopedic knowledge of *E. coli.* I'm so grateful that Debby came to the lab for part of her sabbatical.

My family is very special to me. Without my parents continued support and love, I would not be writing this section. I must also thank my sister, Angie, who almost never picks up the phone, but always has to drive. Also, even though he's only two, my nephew, because he always makes me laugh. I love you all.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

OVERVIEW

For my graduate studies, I have been working with a family of transcriptional regulators called the LysR-Type Transcriptional Regulators (LTTRs) as a model to study protein-protein interactions. In this dissertation, I describe the work that I have done to understand the oligomeric properties of the LTTRs. In this first chapter, I will first introduce the basic concepts of protein oligomerization. Second, I will examine general characteristics of protein-protein interfaces. Third, I will review the LTTRs from *E. coli* K-12 that have been studied. Finally, a review of the LTTRs with available high-resolution crystal structures is provided.

PROTEIN-PROTEIN INTERACTIONS

Protein-protein interactions regulate and drive biological processes in highly specific manners and they consist of polypeptide chains associating to form a particular quaternary structure. These polypeptide chains can be encoded by one or more genes. If the polypeptides are identical, the complex is said to be homotypic, while two or more different polypeptide chains interacting is said to be heterotypic. Larger complexes can contain both homo- and heterotypic interacting polypeptides, but by definition, the minimum number of interacting polypeptides is two. However, the cap on the number

---

This dissertation follows the style of *Protein Science.*

of interactions is essentially unlimited.

Protein-protein complexes are abundant in many biological processes. A review of the enzymes involved in glycolysis show that all enzymes except phosphoglycerate kinase are comprised of multiple subunits (Klotz et al. 1975). Furthermore, all enzymes in the TCA cycle have a quaternary structure that is higher than a monomer. What advantage, though, does a cell gain by using protein complexes instead of one large polypeptide? Several advantages have been proposed. First, utilizing smaller subunits to assemble into larger complexes can require less DNA, allowing for the size of the genome to be smaller than what would be needed for processing many single, large macromolecules. Secondly, a smaller polypeptide restricts the number of errors introduced during the processing of the polypeptide than a larger polypeptide. Finally, genes encoding an oligomer with either an advantageous or deleterious mutation are more rapidly selected for or against than a monomeric protein, most likely due to the higher copy number of the subunits in the complex than that of a monomeric protein (Klotz et al. 1975).

Recent studies of protein-protein interactions have focused on cataloguing those proteins that form complexes and characterizing the physical and chemical nature of the interior interface of an oligomer; it is the latter on which I will focus.

CHARACTERISTICS OF THE PROTEIN-PROTEIN INTERACTION INTERFACE

The physical and chemical properties of a protein-protein interaction are important as they contribute to the specificity and stability of the oligomer. Studies analyzing a set of protein-protein interactions are comprised of protein complexes crystallized and deposited in the Protein Database (PDB) (Berman et al. 2000). The dataset are non-redundant within themselves and do not contain structures that fall into domain-swapping categories (Liu and Eisenberg 2002). Domain-swapping proteins will not be discussed further. Factors such as the size of the interface, the type and number of bond interactions, the types of residues involved in the interface and the contribution of these residues to oligomerization have been examined (Argos 1988; Janin et al. 1988; Jones and Thornton 1995; 1996; Chakrabarti and Janin 2002; Wodak and Janin 2002; Bahadur et al. 2003). These factors and their influence on protein-protein interactions are discussed below.

Types

Jones and Thornton (1996) categorized protein-protein complexes into permanent and non-obligate complexes. By definition, the subunits that make the permanent complexes exist only in the complex, meaning that these proteins are not found as monomeric subunits without denaturation in vitro. Though homooligomers contribute largely to this class, hetero complexes are also found. Non-obligate complexes can exist independently as individual subunits and assemble into the

appropriate protein-protein complexes. Primary examples include enzyme-inhibitor and antibody-antigen complexes. Specifically, Jones and Thornton (2000) give cytochrome *c* peroxidase (CCP) and cytochrome *c* in yeast as examples of non-obligate association. Both exist separately, but interact with one another for respiration in yeast.

Size

The size of the interface can be measured by calculating the change in solvent accessible surface area (ASA) when going from a monomeric to the oligomeric state (Chothia 1974). Upon complex formation, part of a monomer of a protein-protein complex is buried in the interface, protecting a portion of the subunit from the aqueous environment. Those residues that are buried are said to be part of the oligomer interface.

The size of the interface in homooligomeric proteins (permanent complexes) is correlated to the size of the protomer (Jones and Thornton 1996). Burial of 5% of the monomer was suggested as a minimum for this stabilization (Argos 1988). In other complexes, such as an inhibitor-enzyme complex, the size of the interface is limited by the size of the smallest participant.

Amino Acid Composition

The types of residues found at the protein-protein interface have been investigated. Jones and Thornton (1995) calculated that the interface was composed of residues that were 47% hydrophobic, 37% polar and 22% charged. The exterior of the protein was comprised of residues that were 41% hydrophobic, 29% polar and 30% charged. The amino acid composition of the interior of the protein was calculated and

determined to be 71% hydrophobic, 23% polar and 6% charged. Comparing the numbers, the interface of the protein-protein interaction is more similar to the exterior of the protein than that of the interior. This observation was also noted by Argos (19888).

Jones and Thornton (2000) compared the residue composition of permanent dimer complexes and non-obligate dimer complexes. They concluded that non-obligate complexes tend to be more hydrophilic as the individual subunits exist in the cell independently, while the permanent complexes are more hydrophobic. Wodak and Janin (2002) also recognize this general conclusion.

The residue frequency of each residue type in the interface has been calculated. In Janin et al. (1988), leucine was the most common residue in the interface, followed by arginine.. Jones and Thornton (1995) agreed with this observation. In their comparison of permanent and non-obligate interfaces, they also observed that histidine, tyrosine and tryptophan had a higher propensity of being in the non-obligate interface (Jones and Thornton 2000).

Bonds

The variety of bonds in a protein-protein interface contribute to the overall stability of the complex and also places restrictions on the structures, forcing the binding faces of an oligomer to be somewhat complementary (Chothia and Janin 1975).

Hydrogen bonds have been suggested to provide specificity of the protein-protein complex. In their study of 75 complexes, Lo Conte et al. (1999) determined that there are average of 10.1 hydrogen bonds per interface. Further, the different types of

hydrogen bonds were calculated: 0.24 were main-chain-main-chain, 0.4 were main-chain-side-chain and 0.36 were side-chain-side-chain.

The number of salt bridge interactions was found to vary among complexes, with a range of zero to five. There is not a clear correlation between the number of salt bridges at an interface and the size of an interface (Jones and Thornton 1996).

Cores and Rims

Because of the differences observed between the obligate and non-obligate residue compositions, Chakrabarti and Janin (2002) observed that it was difficult to identify protein interaction sites based on chemical composition. Thus, they took a different approach in studying protein-protein interfaces.

Using 70 non-obligate complexes, they demonstrated that the interfaces can be subdivided into different patches. In the larger interfaces (>2000 $Å^2$), there were several different patches. In interfaces < 2000$Å^2$, there was one patch identified. Within each patch, a core and rim is identified. A core residue is defined as having at least one atom of a residue buried in the interface. A residue is in the rim if they contain only accessible atoms. In a study of 122 homodimers, Bahadur et al. (2003) determined that the core had an amino acid composition similar to that of the protein interior. The core of the patch was more hydrophobic. In approaching the rim, the hydrophobicity decreased.

CONTRIBUTIONS OF INDIVIDUAL RESIDUES

While general characteristics of protein-protein interactions can be described, these properties give little insight into the energetic contributions of individual residues.

Determining which residue in the interface contributes to the binding of a complex yields important information concerning complex formation. Replacing a residue with an alanine generates a hole in place of atoms distal to Cß in the wild-type residue. Studying protein-protein interactions using alanine scanning is useful because it allows for the selective removal of individual side chains to probe the contribution of that specific side chain to protein binding. The hole created by the alanine can do one of several things to the complex. The bond between the subunits may be weakened, decreasing the affinity of the monomers. Conversely, the bond may have been a de-stabilizing contributor to the complex. The removal of the wild-type residue thus may stabilize the complex. Finally, replacing the residue may do nothing to the complex.

These ideas were most clearly demonstrated in examining the human growth hormone (hGH) receptor binding with the extracellular domain of hGHbp (Clackson and Wells 1995). In this study, thirty-three alanine mutations were made at the interface of the hGHbp. Only two residues had a large change in binding energy (>4.5kcal/mol), while significant effects were seen for four residues (1.5-3.0 kcal/mol). Further, these residues were in contact with other residues that did not contribute greatly to the binding free energy.

Residues that contribute significantly to the overall binding energy of a protein-protein complex are termed hot spots. ASEdb, the first database to have assembled alanine-scanning data, was assembled to anaylse the available data as a whole, identifying several trends (Bogan and Thorn 1998). First, despite the large size of binding interfaces, single residues contribute a large fraction of the free-binding energy

in an interface. Second, these hotspots are generally surrounded by energetically unimportant contacts referred to as an O-ring around the hotspots. Third, hot spots are generally near the center of the interface and in contact with other hot spots of the opposite subunit. In addition to the Bogan and Thorn dataset, Hu et al. (2000) found that hot spots are enriched for tryptophan, tyrosine and arginine.

Alanine-scanning continues to be an important technique in studying the contribution of a residue to the oligomerization of a protein and new databases continue to attempt to mine the literature. Contributions to AESdb are still accepted. However, while submissions of data are still accepted, active curation of the current literature does not appear to be a priority. Ninety-one protein-protein interaction complexes are noted, representing 118 different proteins. Fischer and colleagues have organized a new database, termed BID (Binding Interface Database) that has mined and curated the primary literature (Fischer et al. 2003). A wiki format has been implemented to allow for community annotation of the interactions. Currently, 467 interaction pairs with 7000 hot spots are in the database (Tsai 2008).

THE LysR-Type TRANSCRIPTIONAL REGULATORS

The LysR-Type Transcriptional Regulators (LTTRs) are a diverse family of oligomeric transcription factors that are found in prokaryotes. The family was initially identified by Henikoff in 1988 with roughly 50 members (Henikoff et al. 1988). With the continuous sequencing of new genomes, that number has grown to over 18,000 potential members (IPR000847 HTH_LysR (Quevillon et al. 2005)), making it perhaps

the largest family of transcriptional regulators among prokaryotes.

In *E.coli* K-12 the target genes of the LTTRs include those that are involved in general functions of nitrogen source utilization, amino acid biosynthesis and catabolism, oxidative stress response and detoxification of the cell (Christman et al. 1989; Schell et al. 1990). In other organisms, however, LTTRs serve a more glamorous role. RovA of *Yersinia pestis* is a global transcription factor that plays a role in virulence (Cathelyn et al. 2006). Work from the Rahme lab has shown MvfR from *Pseudomonas aeruginosa* PA14 serves a critical role in pathogenicity by regulating quorum-sensing regulated virulence factors (Xiao et al. 2006a; Xiao et al. 2006b). In *Vibrio cholerae,* AphB, in cooperation with AphA, is needed for activation of the ToxR virulence cascade, by activating the *tcpPH* operon (Kovacikova and Skorupski 1999).

In general, LTTRs act as transcriptional regulators, typically activating their target gene or operon and negatively regulating their own gene. Most are homotetramers and respond to the presence of a small-molecule inducer. The regulators are able to bind to DNA in the absence of the small-molecule. In their presence, however, there is a conformational change that alters the binding and/or bending of the DNA. They have a general binding site of T_2-N_{11}-A_2. Table 1.1 shows a summary of the 46 *E. coli* K-12 LTTRs as identified by version 1of the COGs database. Below, I will discuss the 20 *E. coli* K-12 LTTRs described in the literature. I have grouped them into three categories: well-characterized, characterized and annotated.

Well-characterized LTTRs

1. AaeR

AaeR is the transcriptional regulator for the *aae* operon and lies upstream of the *aae* promote. The operon was identified using DNA microarrays by exposing *E. coli* cells to *p*-hydroxybenzoic acid (pHBA) (Van Dyk et al. 2004). The three genes in the *aae* operon had at least a ten-fold increase in transcript level in the presence of pHBA. AaeA and AaeB comprise an efflux pump and membrane protein, while AaeX is a small protein of unknown function.

AaeR was demonstrated to be necessary for the expression of the operon by introducing a plasmid containing the *aaeXAP-luxCDABE* gene fusion into a strain containing a null mutation in *aaeR*, as well as the isogenic parental strain. When both strains were exposed to various levels of pHBA, there was up to a 90-fold increase of luciferase activity in the wild-type as compared to the null mutant. pHBA, which is an intermediate in ubiquinone biosynthesis, plays the role of a signal sensor for metabolic imbalance in the cell and is thought to be the inducer of AaeR, though specific binding of pHBA to AaeR has not been demonstrated. The concentration of pHBA is normally

Table 1.1. Summary of the *E. coli* K-12 LysR-Type Transcriptional Regulators

	Member	**Other Name**	**Target Gene, Operon or Regulon**	**Associated Small-molecule***
1.	AaeR	yhcS	*yhcS*RQP/ *aae*ARXAB	*p*HBA, salicylate, benzoate,
2.	Alls	ybbs, glxA1	*allD*	allantoin or glyoxylate
3.	ArgP	iciA	*argK, argO, dnaA*	L-arginine
4.	Cbl		cysteine regulon	5'-phosphosulphate
5.	CynR		*cyn* operon	Cyanate
6.	CysB		cysteine regulon	*O*-accetyle serine; *N*-accetyle serine
7.	DsdC		*dsdA*	D-serine
8.	GcvA		*gcv* operon	Purines and glycine
9.	HcaR	phdR, yfhT	*hca* operon	3-phenylpropionic acid
10.	IlvY		positive regulator of ilvC	Acetoacetate, acetohydroxyburate
11.	MetR		*metA, metE, metH, glyA*	Homocysteine, methionine
12.	NhaR	antO	nhaA, *pgaABCD*	Na^+ or Li^+
13.	OxyR	momR	activator of *oxyS*	Oxidative stress response
14.	TdcA		*Tdc* operon	None known
15.	XapR		xapA	Xanthosine
16.	AbgR	ydaK	Possibly *abg* operon	*p*-aminobenzoyl-glutamate
17.	LeuO		*Leu* operon	None known
18.	LrhA	genR	flagellar, motility and chemotaxis genes	None known
19.	LysR		positive regulator of *lys*	Diaminopimelate
20.	Nac		nitrogen assimilation control proteins	None known
21.	YafC		putative LysR-type transcriptional regulator	Not known
22.	YagP		putative LysR-type transcriptional regulator	Not known
23.	YahB		putative LysR-type transcriptional regulator	Not known

Table 1.1. Continued

	Member	Other Name	Target Gene, Operon or Regulon	Associated Small-molecule*
24.	YbdO		putative LysR-type transcriptional regulator	Not known
25.	YbeF		putative LysR-type transcriptional regulator	Not known
26.	YbhD		putative LysR-type transcriptional regulator	Not known
27.	YcaN		putative LysR-type transcriptional regulator	Not known
28.	YcjZ		putative LysR-type transcriptional regulator	Not known
29.	YdaK		putative LysR-type transcriptional regulator	Not known
30.	YdcI		putative LysR-type transcriptional regulator	Not known
31.	YdhB		putative LysR-type transcriptional regulator	Not known
32.	YeaT		putative LysR-type transcriptional regulator	Not known
33.	YeeY		putative LysR-type transcriptional regulator	Not known
34.	YeiE		putative LysR-type transcriptional regulator	Not known
35.	YfeR		putative LysR-type transcriptional regulator	Not known
36.	YfiE		putative LysR-type transcriptional regulator	Not known
37.	YgfI		putative LysR-type transcriptional regulator	Not known
38	YgiP		putative LysR-type transcriptional regulator	Not known
39.	YhaJ		putative LysR-type transcriptional regulator	Not known
40.	YhjC		putative LysR-type transcriptional regulator	Not known
41.	YiaU		putative LysR-type transcriptional regulator	Not known
42.	YidZ		putative LysR-type transcriptional regulator	Not known
43.	YifD+A		putative LysR-type transcriptional regulator	Not known
44.	YjiE		putative LysR-type transcriptional regulator	Not known
45.	YneJ		putative LysR-type transcriptional regulator	Not known
46.	YnfL		putative LysR-type transcriptional regulator	Not known

*This does not mean binding of the molecule to the protein. The small-molecule merely influences the activity of the LTTR.

low. When metabolic processes are unbalanced, the level of pHBA increases, allowing for the activation of the *aae* operon.

2. AllS

The allantoin regulon consists of genes that are transcribed from three promoters: *allA*, *glc* and *allD* (Cusa et al. 1999). Under anaerobic conditions, the genes in the regulon allow for the utilization of allantoin as a sole source of nitrogen, breaking down the allantoin into ureidoglycolate. The ureidoglycolate has two fates: $NH4^+$, CO_2 and ATP or 3-phosphoglycerate. Allantoin and glyoxylate induce all three promoters; however, different gene products are required depending on the condition. *allA* and *glc* are expressed under both aerobic and anaerobic conditions, while *allD* is only expressed under the anaerobic conditions. AllR represses *allA, glc* and *allS* in the absence of glyoxylate.

allS, also known as *ybbS*, encodes a transcriptional regulator that targets *allD* (Rintoul et al. 2002). Binding of AllS to the *allD* promoter was shown to be independent of any intermediates. Further, the binding site of AllS had the LTTR binding motif of T_5-N_5-A_2-C-A_2.

3. ArgP

Mutations in *argP* were identified in cells that were able to grow in the presence of canavanine, which is a natural analog for arginine. It was observed that there was a 40% reduction in the transport of L-arginine, L-lysine and L-ornithine. Additionally, the

ATP binding protein ArgK had a 20% reduction in activity. The mutations in these canavanine resistant strains mapped to *argP*. Thus, the author concluded ArgP, the product of *argP,* is a regulator of the process of transporting arginine (Celis 1999). In vitro transcription assays showed that ArgP was able to increase transcription of *argK* seven-fold. The addition of 10mM L-arginine inhibited binding of ArgP to the DNA.

In the same study, Celis observed that the previously published sequence called *iciA* was identical to that of *argP* (Thony et al. 1991). IciA was previously identified as an inhibitor of chromosomal initiation replication in vitro (Hwang et al. 1992) as well as an activator of transcription of *dnaA* (Lee et al. 1996; Lee et al. 1997). The author proposes that ArgP and IciA are the same protein (Celis 1999).

4. Cbl

Cbl, standing for CysB-like, is responsible for the regulation of the *tau* and *ssu* gene clusters. These genes allow for the utilization of taurine or alpiphatic sulphonates for the liberation of sulphite, which enters the cysteine biosynthesis system. These genes are repressed under conditions with cysteine or sulphate present. Cbl and CysB were shown to bind directly to the *tau* and *ssu* promoters (van Der Ploeg et al. 1999). Cbl and not CysB is only able to activate transcription, with CysB having a negative effect on the *ssu* promoter (Bykowski et al. 2002). Expression from *ssu* promoter by Cbl was inhibited by 5'-phosphosulphate (APS) (van Der Ploeg et al. 1999). *cbl* is activated by CysB, the master regulator of sulphur assimilation (see below).

5. CynR

CynR is the transcriptional regulator of the *cyn* operon, which encodes genes that allow cyanate to be used as a sole source of nitrogen. The operon consists of *cynT, cynS, cynX*, which encode a carbonic anhydrase, a cyanase and a protein of unkown function, respectively (Sung and Fuchs 1988). Plasmids encoding for *cynR* were able to restore the function of CynR in strains deleted for *cynR* (Sung and Fuchs 1992). In DNA binding experiments in vitro, CynR was able to bind to DNA in the presence or absence of cyanate. However, CynR binding induced bending of the DNA. The amount of DNA bending was decreased in the presence of cyanate (Lamblin and Fuchs 1994).

6. CysB

CysB is the positive regulator of the *cys* genes, which are needed for the uptake and reduction of oxidized forms of inorganic sulfur to sulfide. Full expression of the genes requires the small molecule inducer N-acetylserine and low sulphur conditions. CysB, like other LTTRs, repress its own transcription (Jagura-Burdzy and Hulanicka 1981). In solution, CysB is a tetramer of four individual subunits. Through the use of circular DNA fragments to study the amount of bending, Hryniewicz and Kredich (Hryniewicz and Kredich 1994) calculated bend angles of 102° and 96° with the *cysK* and *cysP* promoters. Upon the addition of N-acetylserine, the bending relaxed to approximately 50° for both promoters. At *cysB*, acetylserine reduces CysB binding to DNA and thus relieves the self-repression.

7. DsdC

DsdC was shown to be a specific activator of *dsdX* and *dsdA,* requiring D-serine for activation. D-serine deaminase, the product of *dsdA*, enables *E. coli* to use D-serine both a sole source of carbon and nitrogen in minimal media, by breaking it down to pyruvate and ammonia; the function of *dsdX* is not known. Otherwise, D-serine is toxic to the cell, as it inhibits L-serine and panthothenate synthesis.

8. GcvA

GcvA is one of five regulators of the *gcv* operon (Heil et al. 2002). This operon contributes enzymes that are part of the glycine cleavage pathway. In the presence of glycine, GcvA activates the *gcv* operon. In the presence of purines a five-fold repression of the operon is observed (Jourdan and Stauffer 1998). Glycine, however, does not bind to GcvA, interacting instead with GcvR. With DNaseI footprinting assays Wilson et al. (Wilson et al. 1995) demonstrated the presence of three GcvA binding sites in the region between *gcv* and *gcvA*. All three sites were needed for repression of *gcv* in the presence of purines, while only two were necessary for activation by glycine.

In their model, Heil et al. proposed that in the absence of glycine, Lrp and GcvA helped to bend the DNA, such that GcvA and GcvR could interact to form a GcvA-GcvR-GcvA complex, with each protein binding at individual sites. In the presence of glycine, the GcvR is displaced, allowing for interactions between the two units of GcvA and the a subunit of the RNA polymerase (Jourdan and Stauffer 1999; Ghrist et al. 2001; Heil et al. 2002).

9. HcaR

HcaR controls the expression of *hcaE, hcaF, hcaC, hcaB* and *hcaD*, which are genes that code for dioxygen detoxification processes, as well as catabolism of 3-phenylpropionic acid. In a 2D gel electrophoresis study comparing wild-type *E. coli* to an *E.coli hcaR::cat* strain, 51 proteins were found whose regulation is influenced by the absence of *hcaR* (Turlin et al. 2005). Expression of several glycolysis and TCA cycle enzymes decreased, while some of the gene products involved in gluconeogenesis increased. Additionally, genes involved in the oxidative stress response where identified: thioredoxin reductase, DNaK and MnSOD. Further investigation demonstrated these observations to be related to the regulatory effect that HcaR has on *hca*A.

10. IlvY

IlvY is the transcriptional regulator for *ilvC*, whose product is acetohydroxy acid isomeroreductase (EC 1.1.1.86), an enzyme involved in the synthesis of branched chain amino acids. The induction of *ilvC* requires IlvY, as well as either of its two substrates: a-acetolactate or a-acetohydroxybutyrate. Binding of IlvY to the *ilvC* promoter was shown to be required, but not sufficient, for the transcription of *ilvC* (Rhee et al. 1998). The presence of an inducer molecule prompted a conformational change in the IlvY-DNA complex, enhancing the RNA polymerase recruitment to the *ilvC* promoter region. This mechanism of the presence of a small-molecule causing a conformational change in the protein-DNA complex is the most common mode of activation in the LTTR family.

The mechanism by which IlvY forms protein-DNA complexes was intensely studied and was the first of the LTTRs to have its mechanism elucidated.

11. MetR

MetR plays a role in the regulation of *met* gene expression. Currently, MetJ and MetR are known regulators of these genes. While MetJ, which is not a LTTR, appears to be the global regulator, MetR is specifically necessary for the expression of *metE,* with homocysteine required for full *metE* expression. Further, the full-expression of *metH* requires MetR, though basal levels of *metH* are higher than that of *metE* in the absence of MetR protein, with homocysteine having a negative effect on *metH* expression. Expression of *metA* was shown to require MetR, though the presence of homocysteine decreased the activation of *metA* by MetR (Mares et al. 1992). Additionally, MetR has shown to play a role in the expression of *glyA*. Plamann and Stauffer demonstrated that MetR and homocysteine were required for activation of *glyA* (Plamann and Stauffer 1989), while Lorenz and Stauffer investigated the binding region of MetR in the upstream region of *glyA* (Lorenz and Stauffer 1995).

12. NhaR

nhaA encodes a sodium antiporter and is responsible for growth in alkaline pH in the presence of Na^+ (Padan et al. 1989). It was demonstrated that a multi-copy plasmid containing *nhaR* could enhance the activation of a *nhaA'-'lacZ* fusion in a Na^+ dependent manner. The effect of NhaR was not seen without Na^+ (Rahav-Manor et al.

1992). Additional studies by Goller et al. showed that NhaR could activate the *pgaABCD* operon. This operon is needed for the production of poly-ß-1,6-*N*-acetyl-D-glucosamine (PGA), which is a necessary molecule in biofilm formation. Strains deleted for *nhaR* had a decrease in the formation of biofilm and PGA was undectable (Goller et al. 2006).

13. OxyR

The transcription factor OxyR senses H_2O_2. OxyR targets genes that are involved in the cells response to oxidative stress, including *oxyS, ahpCF, katG, dps, sufA, grxA, gorA* and itself, *oxyR*. Two positions, Cys199 and Cys208 were shown to be important for the activation of OxyR. Changing these positions to serine caused OxyR to be unable to activate its target genes. Upon exposure to H_2O_2, there is a large subunit rotation (as shown in the crystal structures (Choi et al. 2001), see below for more details) to facilitate bringing these two cysteines closer together and allowing for disulfide bond formation. The redox potential of OxyR was determined to be -185mV, well below the cells internal redox potential of -260mV (Zheng et al. 1998). Therefore, OxyR is reduced under normal growth conditions.

14. TdcA

TdcA is an unusual LTTR, as it is transcribed with the genes that it targets: the *tdc* operon. The *tdc* operon is involved in the transport and metabolism of threonine and serine during anaerobic growth. It is interesting to note that there is an upstream,

divergently transcribed gene, *tdcR*, which is necessary for the expression of *tdc*. There are two mechanisms of *tdc* expression: one that is influenced by metabolic intermediates and one that is under anaerobic control. The latter is described elsewhere (Chattopadhyay et al. 1997; Sawers 2001). The former is influenced by CAP and IHF, with TdcR and TdcA required for full expression. *tdcA* was identified as being a member of the LTTRs, as well as demonstrating that the *tdcA* gene product could act in *trans* as a positive regulator of *tdc* (Ganduri et al. 1993). Further, work from the Datta lab went on to show that all named factors above act together for the regulation of *tdc,* suggesting bending and looping of DNA were involved (Hagewood et al. 1994).

15. XapR

In *E. coli,* xanthosine is cleaved into the nucleoside base and pentose-1-phosphate in a reaction catalyzed by xanthosine phosphorylase, which is encoded by *xapA* (Seeger et al. 1995). *xap*A lies in an operon with *xapB*, which encodes a membrane protein similar to NupC, which is a nucleoside transport protein. *xapA* and *xapB* expression is dependent on the transcriptional regulator, XapR (encoded by the gene *xapR*), as well as xanthosine.

XapR, unlike the majority of LTTRs, is constitutively expressed and not autoregulated. Mutant studies of XapR have identified regions important for xanthosine binding. Additionally, these studies have identified mutants that can be induced by other nucleosides; in particular, deoxyinosine (Jorgensen and Dandanell 1999).

Characterized

16. AbgR

AbgR is implicated as the regulator of the regulator of *abgA, abgB* and *abgT* genes whose products are involved in the uptake and catabolism of *p*-aminobenzoate. AbgR is transcribed in the opposite direct of *abg*. AbgR has 27% identity to TdcA (see below) (Hussein et al. 1998). They also found in the *abgR-abgA* intergenic region two half sites of 5' –GATAA-3', which is one-half of the T-N_{11}-A typical LTTR binding site (Carter et al. 2007). Taken together, this is strong evidence that AbgR is the regulator of *abg;* however, further investigation is needed.

17. LeuO

LeuO is a complicated LTTR. It has been shown to have several jobs in *E. coli.* First, it has been shown to activate the *leuLABCD* operon. Second, it has been shown to affect the expression of DsrA-RNA, which is a small regulatory RNA (Klauck et al. 1997). Through this regulation, LeuO indirectly reduces *rpoS* levels at low temperatures. Additionally, LeuO has been shown to relieve H-NS dependent *bgl* silencing. The expression of LeuO is sensitive to levels of ppGpp (Fang et al. 2000). H-NS severely represses the expression of *leuO*.

18. LrhA

LrhA was identified as a global regulator for chemotaxis, flagella and motility. Comparing the expression of MG1655 and a MG1655 *lrhA*$^{-}$ identified a large number of genes whose expression was upregulated when compared to the wild-type MG1655. In the same study, it was further shown that, unlike the majority of LTTRs, LrhA positively regulates its own expression. There is no known inducer of LrhA.

19. LysR

LysR is the namesake of the LTTR family. It regulates *lysA,* which encodes a diaminiopimelate decarboxylase and is responsible for the conversion of diaminopimelate to lysine (Stragier et al. 1983). It was originally known that mutations in *lysA* led to a Lys^{-} phenotype. Attempts to more precisely map the *lysA* gene led to the discovery of a second class of mutants that led to the Lys- phenotype: mutations that were found in *lysR*. Stragier and colleagues went on to further characterize LysR to identify the characteristic divergently transcribed promoters of the LTTR to its target gene (Stragier and Patte 1983).

20. Nac

The nitrogen assimilation control protein, or Nac, appears to serve a role of transcriptional regulator in nitrogen metabolism of *E. coli.* Among the genes it targets, it represses *gdh* and *asnC,* while activating *codAB*, *nupC* and *gabDTPC* (Muse and Bender 1998; Zimmer et al. 2000; Poggio et al. 2002; Muse et al. 2003). There is no clear

inducer of *nac*. It should be noted that Rosario and Bender (Rosario and Bender 2005) observed Nac from *Klebsiella pneumoniae* to be a tetramer while Nac from *E. coli* appears to be a dimer.

Annotated

The following gene products have been annotated based on the clusters of orthogolous groups (Tatusov et al. 1997) : YafC, YagP, YahB, YbdO, YbeF, YbhD, YcaN, YcjZ, YdaK, YdcI, YdhB, YeaT, YeeY, YeiE, YfeR, YfiE, YgfI, YgiP, YhaJ, YhjC, YiaU, YidZ, YifD+A, YjiE, YneJ, YnfL. At the time of this writing, no further information could be found on these gene products.

ORGANIZATION OF THE LTTR POLYPEPTIDE

The LTTRs are relatively the same size of roughly 300 residues. Numerous mutational experiments were done to probe the mechanism of transcriptional activation by LTTRs, identifying several regions with different functions. The N-terminus, consisting roughly of the first 60 residues, is responsible for the binding of the DNA molecule. Several studies have identified residues in this region that are necessary for the activation of the target gene. Residues 92- 271 comprise the inducer binding/oligomerization domain. Finally, residues 272-298 contribute to the activation of the target gene and, in some cases, to the DNA binding. This organization is shown in Table 1.2.

Table 1.2. Organization of the Polypeptide of the LTTR

DNA Binding Domain	Helical linker	Oligomerization/ Inducer Binding Domain	Activation
1-60	59-88	92-271	272-298

The N-terminal DNA binding domain was shown to consist of a helix-turn-helix motif. Though previously published similarity scores only included about 50 LTTRs, there was a >20% identity in this region amongst the LTTRs (Schell 1993). Mutational analyses of several LTTRs have probed the role of residues in the DNA-binding domain to identify two main classes of mutations. The first class consists of mutations that contribute to the DNA-binding of the protein. GcvA from *E. coli* requires positions V32 and S38 for binding DNA (Jourdan and Stauffer 1998). Mutagenesis studies in OxyR found that the following mutations: R4C, Y8C, A22V, P30L, T31M, L32F, S33N and R50W significantly lowered or even abolished the ability of OxyR to activate transcription at P_{oxyS} (Kullik et al. 1995a; Wang et al. 2006). In CysB several mutations were found to affect DNA binding: I33N (Colyer and Kredich 1994), S34R (Kredich 1992), T21, E41K, L44R and I48T REF. Further, in NahR from *Pseudomonas* several positions including T26, A27, P35, R43, R45 and T56 had importance for DNA binding. It was thought that, in general, residues 23-45 held significance for protein-DNA interactions (Schell 1993). However, with additional LTTRs being identified, that window could be expanded. The second class of mutations in this region affected

activation. In the same study of GcvA, positions L30 and F31 were identified as causing a defect in activation of *gcvT-lacZ* fusion (Jourdan and Stauffer 1998).

Using the CbnR numbering (because it is the only LTTR with an available high-resolution crystal structure for the full protein), residues 59-88 consist of a region that is comprised of a coil-coil domain. This region serves as a linker, connecting the regulatory domain to the DNA binding domain. A systematic and detailed analysis of this region is not found in the literature.

The C-terminal domain, comprised of residues 88- to the end of the protein is known both as the regulatory domain and the effector binding domain; the majority of oligomerization occurs here. The details of the available structures are discussed elsewhere (see below). However, extensive mutational analysis looking for an inducer binding site have mapped the area to the C-terminal domain, thus the name. The overall structure of this domain looks like a periplasmic sugar binding protein.

CbnR: The Full-length Structure

Although there are many members in the LTTR family, there are only a few high-resolution atomic structures available. The first structure of a full-length LTTR was CbnR from *Ralstonia eutropha* NH9. Ogawa et al. demonstrated that *cbnABCD* was the target of CbnR by utilizing transcriptional fusions in the presence of 3-chloroobenzoate or benzoate, two inducers of CbnR. In the absence of inducers, a 60 bp region from -20 to -80 upstream of *cbnA* was protected by CbnR (DNase I footrpinting). Further, they showed a bending of DNA occurred (70°) without the inducer. In the

presence of *cis-cis* muconate, this was relaxed to 54°. It was noted that the footprint did not change in the presence of 1mM inducer.

The structure of CbnR was solved to a resolution of 2.5Å (Muraoka et al. 2003) and was comprised of a homotetramer, consisting of a dimer of dimers. Examination of the monomers shows two distinct conformations: extended and compact. It is hypothesized that these conformations facilitate the promoter binding and DNA bending, as described above (Muraoka et al. 2003).

The regulatory domain of CbnR can be divided into two subdomains, regulator domains I and II. RDI, consisting of residues 88-161 and 265-294, contains a core of five ß-strands, three a-helices and a 3_{10}-helix. RDII is similar to that of RDI. RD1 and RDII are connected by a ß-strand and a loop. Both features combined is called hinge region 3. It is thought that this hinge region undergoes a conformational change upon inducer binding. Although the inducer was present in the crystallization reaction, it was not found in the structure. However, mutational analysis of CysB, OxyR and NahR strongly suggest that the binding site is located in the hinge region 3. (Schell et al. 1990; Kullik et al. 1995a; Lochowska et al. 2001)f

Structures of Regulatory Domains: BenM, CatR, Cbl, CynR, CysB, DntR and OxyR

Crystallization of full-length LTTRs has been described as being difficult, due to solubility issues. Further, the inducer's ability to bind to the protein may be hampered by the high-salt conditions of crystallization. (Ezezika et al. 2007). Nonetheless, high-

resolution structures of regulatory domains were available before the publication of the CbnR structure, and continue to be described thereafter. The important structural features of the available regulatory domain crystal structures are depicted below.

1. BenM

BenM, from *Acinetobacter baylyi* ADP1, works with CatM for aromatic compound degradation. In the presence of benzoate, BenM works with CatM to activate genes for the utilization of benzoate as a sole-source of carbon and represses genes that consume other aromatic compounds. The combined presence of both *cis,cis*-muconate and benzoate (both effectors of BenM) was shown to have a synergisitic effect on transcriptional activation on *benA* (Bundy et al. 2002).

The structures of BenM and CatM are the only structures available that have physiologically relevant effectors co-crystallized. The effector binding domain was crystallized and found to be an a/ß domain much like that of CbnR (Ezezika et al. 2007). As in other C-terminal domains, there are two regions separated by a hinge region. Domain I consists of residues 87 to 161 and residues 268 to the end of the C-terminus. Domain II is comprised of residues 162-267.

cis,cis-muconate was present in one subunit of BenM between domain I and domain II. This site has been designated the primary binding site and is highly conserved in CatM (see below). A molecule of benzoate was discovered in a secondary effector-binding site. The authors speculate that this site must be a high-affinity site, as exogenous benzoate was not present in crystal conditions.

In crystals that were soaked in benzoate, benzoate was bound to both the primary and secondary effector-binding sites. The benzoate interacts in a mainly hydrophobic fashion, with residues L100, L105, I108, F144, L159, I269 and I289 interacting with the ring of benzoate. Comparison of structures with bound and unbound benzoate in the secondary site led the authors to predict that a charge-relay system enhances the signal from the bound muconate in the primary site. At the heart of their argument is Glu162. When benzoate is not present in the site, Glu162 forms a salt-bridge with Arg160. They speculate that when benzoate binds at the secondary site, Arg160 forms a salt-bridge with the carboxyl of benzoate. This causes Glu162 to form a salt-bridge with Arg146, which is located in the primary site, where *cis,cis*-muconate is bound. Formation of the salt-bridge leads to an increased electrostatic potential on the *cis,cis*-muconate, which allows for the helixes to tighten up on the *cis,cis*-muconate. The authors suspect that this conformational change will then affect transcription (Ezezika et al. 2007).

2. CatM

CatM from *Acinetobacter baylyi* ADP1 works with BenM for the utilization of benzoate as a sole source of carbon. Unlike BenM, CatM only responds to *cis,cis*-muconate as an inducer. BenM and CatM have 59% sequence identity with an overall similarity of 85%. The structure of CatM's effector binding domain was solved by Momany and colleagues (Ezezika et al. 2007).

The overall structure of CatM is similar to that of BenM. The a/ß fold is the same as seen in other C-terminal domains of LTTRs. Both the structures with bound

and unbound *cis,cis*- muconate were solved. The muconate molecule was found in the equivalent primary binding site of BenM and the protein-molecule interactions are similar. Extensive hydrogen bonding between the carboxylates of the *cis, cis*- muconate occurs with residues comprising the effector binding pocket, including Ser99, Thr128, Lys129, Pro201, F203 and Val227. Various water molecules are also involved. A difference between the binding of the *cis,cis*- muconate of BenM and CatM is seen at position 98. In CatM, there is a serine, which does not interact with the molecule. In BenM, position 98 is glycine. In BenM, a water molecule is able to fill the space at the position.

3. Cbl

The regulatory domain of Cbl was published in 2006 by Bujacz and colleagues (Stec et al. 2006). The structure was reported to 2.8 Å resolution and the effector 5'-phosphosulphate was modeled. Residues 88-307 were included in all four chains solved. Two subdomains exist and the crossover region is residues 163-166 and 266-269. RDI has residues 88-162 and 270-307, while RDII has residues 166-265.

4. CynR

The structure of the regulatory domain of CynR with no inducer was deposited into the PDB (2HXR). The authors were kind enough to share the structure of the regulatory domain with the inducer (Personal Communication, Alexi Savchenko). These structures are addressed further in Chapter III.

5. CysB

Tyrell et al. (Tyrrell et al. 1997) described the 1.8Å crystal structure of residues 88-324 of CysB from *Klebsiella aerogenes* (PDB 1AL3). The fragment consists of two a/ß domains, which the authors call I and II. The two regions are connected by cross-over regions encompassing 163-166 and 266-269. Region I contains residues 88-162 and 270-292, while II is comprised of residues 166-165. The first domain having five ß strands, which form a ß-sheet, and four a helices. The second domain is similar to the first, having five ß sheets and only three helices.

A cleft of 10 Å wide with a diameter of 6 Å was found and included residues 100, 101-103, 149, 164, 166, 202 and 270. This was thought to be an inducer binding pocket. The dimerization interface is formed by a_1, a_2 and $ß_B$, (residues 101-106, 109-118 and 121-128, respectively) which pack against a_{VI} and $ß_G$ (227-235 and 219-224, respectively) of the opposite monomer.

6. DntR

DntR was isolated from *Burkholderia* sp. Strain DNT. This strain (in addition to several others) can utilize the synthetic compound 2,4-dinotrotoluene (2,4-DNT) as a sole source of carbon. 2,4-DNT is of interest, as it is used in the manufacturing of pesticides, explosives and munitions, and is highly toxic to animals. DntR was studied with a desire to design a mechanism for the detection and removal of such compounds. The gene for DntR was found upstream of those genes involved in the degredation of

2,4-DNT to usable metabolic intermediates. Though DntR responded poorly to 2, 4-DNT as an inducer, it was extremely sensitive to salicyclic acid. The desire to rationally redesign DntR to better bind 2, 4-DNT led the authors to solve the crystal structure of DntR from *Burkholderia* sp. DNT.

The structure was solved in 2004 (Smirnova et al. 2004). The full-length protein was present, however only the C-terminal domain was able to be refined to a high-resolution. 1,539Å^2 were buried at the interface. The overall structure is similar to other LTTR C-terminal domains, with a head-to-tail orientation. Residues 90-166 and 274-301 make the RDI, while residues 171-269 make up RDII. The hinge region is comprised of residues 167-170 and 270-273. Salt bridge interactions between the monomers were found at H228 and D105 and R248 and E300. Acetate and thiocyanate were found in the inducer binding cavity and this allowed the authors to model in the saliclate and 2,4-dinitrotoluene.

The crystals of another crystallization condition than that from above showed the homotetramer. Though only at 3Å, these crystals had cleaner helix-turn-helix domain electron densities, allowing the authors to generate a full-length model of DntR. The overall structure looks similar to that of CbnR, with monomeric units having either an extended or compact conformation. Superposition of the CbnR and DntR molecules show the helical linkers positioned differently. This positions the HTH differently then in CbnR, putting the HTH farther apart in DntR then in CbnR. The authors believe that because CbnR had no molecule in the binding pocket, while DntR did, the CbnR molecule should be considered an inactive conformation, while the modeled DntR

structure should represent the active conformation (Smirnova et al. 2004).

7. OxyR

OxyR is a transcription factor that senses H_2O_2. The two structures (PDB 1I69 and 1I6A) for OxyR contain the regulatory domain in both a reduced and oxidized form. When exposed to H_2O_2, there is disulfide bond formation between Cys199 and Cys208. The oxidized OxyR can then induce the cooperative binding of RNA polymerase to activate transcription. Similar to the other structures, OxyR consists of two α/β domains that are connected *via* two interdomain strands. Domain I has four β-strands and two α helices with domain II having β-strands 5-10 and helices C and D. There is another β-strand and helix that is at the end of domain II that reaches up to interact with domain I. Cys199 is located between the two domains, while Cys208 is at the lower end of domain II. In the reduced form, these two residues are substantially separated.

The transition from reduced to oxidized form results in a major structural conformational change. There is a large rotation between the two subunits of approximately 30°. In the oxidized form, the two loops for the active cysteines move to allow disulfide bond formation, bringing the Cys199 and Cys208 closer together. The authors of the paper believe that there is a rearrangement of the hydrophobic core between the reduced and oxidized forms. In the reduced form, A233 interacts with I110 and L124 of the opposite monomer, while in the oxidized form, the core is reformed between F219 and the I10 and L124. Many proline residues stack to form a flat helix surface that helps the conformational change between the reduced and oxidized forms.

These include prolines 99, 103, 107 and 111.

SUMMARY

In summary, the global features of the regulatory domains of the LTTRs with available high-resolution atomic structures are the same, as shown in Figure 1.1. However, major differences between the structures are seen in the external faces of the structures. Closer examination of the structural features of the interfaces highlights that the structural motifs are the same. Differences of the interface features, as shown in Figure 1.2, are observed. Though the same motifs can be found at the interface, the positioning of the helix and strand to the opposite monomer show differences. How the residues that comprise these interface structures interact with one another is the global question I seek to address.

OVERVIEW OF DISSERTATION

This dissertation is an effort to dissect the residues that contribute to the oligomerization of the LysR-Type Transcriptional Regulators. In Chapter II, I describe the work that was done to identify residues important for the oligomerization of OxyR. In Chapter III, I describe our efforts to determine the regions of CynR necessary for oligomerization. Further, we predicted the residues identified from OxyR which may identify important residues for the oligomerization of CynR. In Chapter IV, I describe the work that was begun to identify LysR-Type Transcriptional Regulators that could form potential heterotypic interactions. Chapter V describes the work done to show that

the AcrAB multidrug resistant efflux pump is responsible for conferring resistance to methotrexate upon *E. coli* laboratory strains. This study was the result of my rotation project. Chapter VI is a global discussion about the oligomerization of the LTTRs.

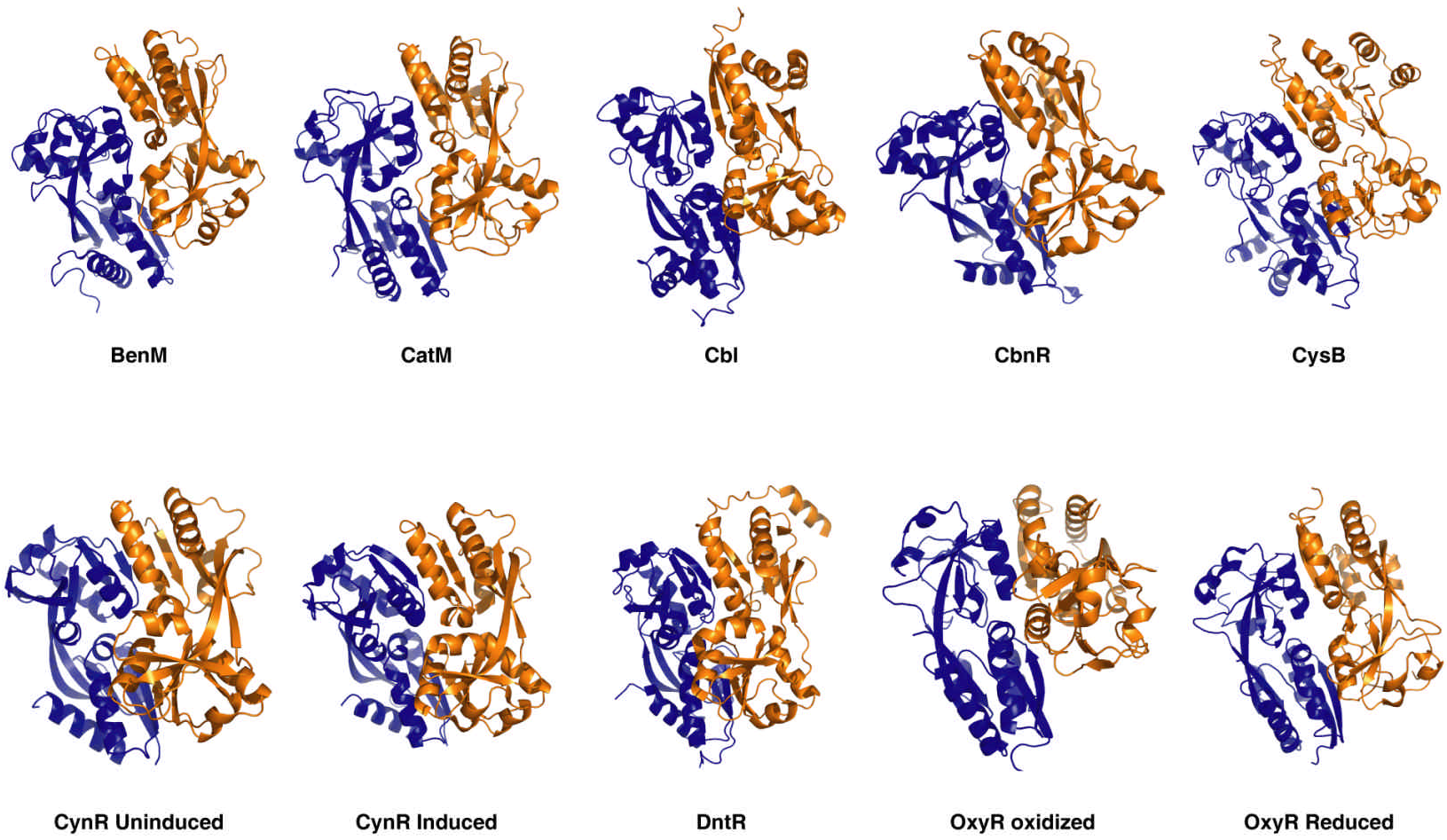


Figure 1.1. Crystal Structures of LTTR Regulatory Domains. One monomer is colored blue, the other in orange.

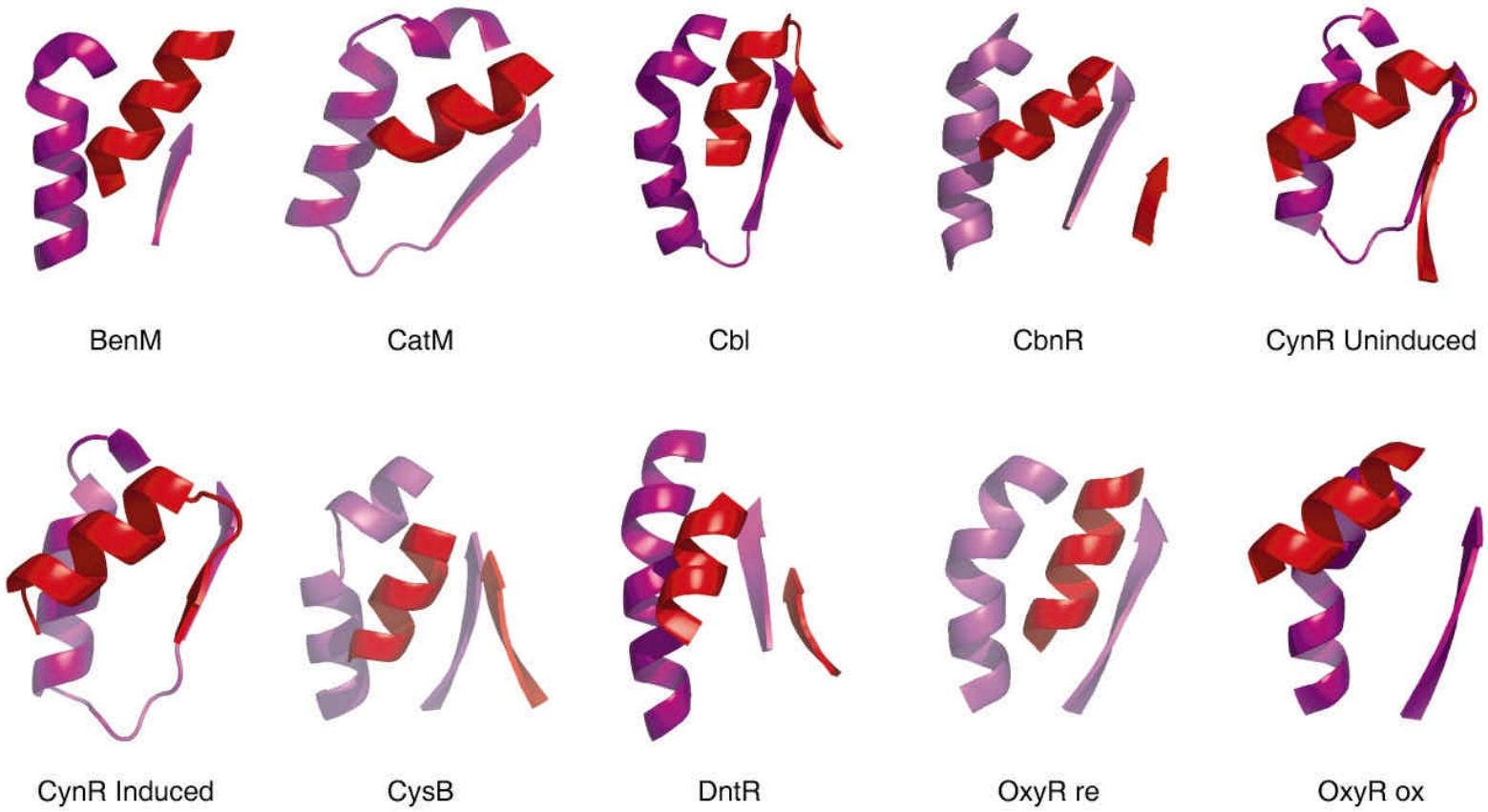


Figure 1.2. The Interface Motifs of the Regulatory Domains. One side of the interface is colored purple, while the other face is colored red.

CHAPTER II

THE OLIGOMERIZATION PROPERTIES OF OxyR IN VIVO

OVERVIEW

We examine the contribution of residues at the dimer interface of the transcriptional regulator OxyR to oligomerization. Residues in contact across the dimer interface of OxyR were identified using the program Quaternary Contacts (QContacts). Alanine-scanning mutagenesis was performed on the non-alanine or glycine residues identified in the resultant contact profile and the oligomerization ability of the mutant proteins was tested using the λcI repressor system to identify residues that are hot spots in OxyR. Interestingly, these important residues for oligomerization are not especially conserved amongst a set of OxyR orthologs. We compare the properties of these hot spots to those described in the literature from other systems.

INTRODUCTION

Protein-protein interactions regulate and drive many biological reactions. Understanding the contributions of individual residues to the oligomerization of a protein is necessary for appreciating the highly-specific nature of the interaction. The LysR-Type Transcriptional protein OxyR is a global response regulator to oxidative stress. In *E. coli*, OxyR controls a regulon including *katG, dps, ahpCF, sufA, grxA, gorA* and *oxyS*. OxyR has been extensively studied to understand the molecular basis of the oxidative stress response in *E. coli*. In its monomeric form, OxyR is a 34kDa protein,

with the active form of OxyR being homoteterameric, consisting of a dimer of dimer. Mutational studies identified Cys199 and Cys208 necessary for the response to oxidative stress (Kullik et al. 1995b).

The crystal structures of the reduced and oxidized forms of the regulatory domain (residues 80-305), solved by Choi et al. (2001) demonstrated a large rearrangement of the subunits, creating two distinct structural forms of the protein characterized by a 30° rotation between the subunits. This rearrangement was brought about by the formation of an intramolecular disulfide bond between Cys199 and Cys208, which are 17Å apart in the reduced form.

Alanine scanning has been used to investigate individual residues' contributions to the oligomerization of many oligomers (Clackson and Wells 1995; Bogan and Thorn 1998). In this report we examine the roles of residues in the subunit interfaces of OxyR in stabilizing the oligomeric forms. We identified residues with potential for contributing to the oligomerization of OxyR utilizing QContacts (Fischer et al. 2006). The generated list of residues in contact, or contact profile, was used as a guide for alanine scanning mutagenesis, and the ability to form oligomers was assessed using the lambda repressor fusion assay. We find a small subset of the candidate residues have strong oligomerization phenotypes in the tetramer, consistent with these residues being hot spots in one or both oligomeric forms of OxyR. Other residues in the interface have weaker but significant energetic contributions to oligomerization, as determined from repressor fusion assays using the regulatory domain instead of the full length OxyR.

METHODS

Generation of the Contact Profile

The crystal structures used in these studies were PDB files 1I69 (reduced) and 1I6A (oxidized) (Choi et al. 2001). The dimeric form of 1I6A was constructed using the PQS Server (Henrick and Thornton 1998). QContacts (Fischer et al. 2006) was used to calculate the residues in contact across each chain. The output of QContacts was filtered for redundant pair-wise interactions and the subsequent culled residue contact profiles are shown in Tables 2.1 and 2.2.

Alanine Scanning Mutagenesis

Using the contact profiles as a guide, the identified non-alanine and glycine residues of OxyR in the λcI OxyR repressor fusion were mutated to alanine. PrimerX (http://www.bioinformatics.org/primerx) was used to generate the primer sequence and primers were ordered from IDT (Iowa). Pfu Turbo (Stratagene) or Pfx Platinum (Invitrogen) was used for the polymerase. pGK702 was the DNA template. Reactions were treated with 2U of *Dpn*I for 2 hours and transformed into Mach1-T1^{R}, Top1-T1^{R} (Invitrogen) or XL-1Blue Supercompetent (Stratagene) cells. Transformants were recovered and sequenced (LPGT, TAMU).

Table 2.1. Identified Residues in Contact in the Oxidized Form of OxyR

Chain	Residue #	Name	Contacts	Chain	Residue #	Name
A	97	LEU	to	B	222	THR
A	102	GLY	to	B	226	THR
A	103	PRO	to	B	226	THR
A	103	PRO	to	B	229	ASN
A	103	PRO	to	B	225	GLU
A	103	PRO	to	B	223	SER
A	104	TYR	to	B	229	ASN
A	104	TYR	to	B	225	GLU
A	106	LEU	to	B	226	THR
A	106	LEU	to	B	230	MET
A	106	LEU	to	B	222	THR
A	106	LEU	to	B	219	PHE
A	107	PRO	to	B	229	ASN
A	107	PRO	to	B	230	MET
A	107	PRO	to	B	233	ALA
A	107	PRO	to	B	226	THR
A	110	ILE	to	B	230	MET
A	110	ILE	to	B	219	PHE
A	110	ILE	to	B	235	SER
A	111	PRO	to	B	233	ALA
A	111	PRO	to	B	230	MET
A	111	PRO	to	B	235	SER
A	114	HIS	to	B	235	SER
A	114	HIS	to	B	219	PHE
A	121	GLU	to	B	218	HIS
A	122	MET	to	B	218	HIS
A	123	TYR	to	B	218	HIS
A	123	TYR	to	B	217	THR
A	123	TYR	to	B	220	ARG
A	124	LEU	to	B	219	PHE
A	124	LEU	to	B	222	THR
A	124	LEU	to	B	218	HIS
A	124	LEU	to	B	221	ALA
A	124	LEU	to	B	220	ARG
A	125	HIS	to	B	221	ALA
A	125	HIS	to	B	222	THR
A	126	GLU	to	B	222	THR
A	126	GLU	to	B	221	ALA
A	126	GLU	to	B	223	SER
A	126	GLU	to	B	226	THR
A	126	GLU	to	B	128	GLN
A	131	GLN	to	B	131	GLN
A	225	GLU	to	B	225	GLU
A	247	PRO	to	B	252	ASP
A	248	GLU	to	B	252	ASP
A	251	ARG	to	B	251	ARG
A	251	ARG	to	B	252	ASP

Table 2.2. Identified Residues in Contact in the Reduced Form of OxyR

Chain	Residue #	Name	Contacts	Chain	Residue #	Name
A	94	HIS	to	B	219	PHE
A	103	PRO	to	B	225	GLU
A	103	PRO	to	B	229	ASN
A	106	LEU	to	B	229	ASN
A	106	LEU	to	B	226	THR
A	106	LEU	to	B	225	GLU
A	107	PRO	to	B	252	ASP
A	107	PRO	to	B	229	ASN
A	107	PRO	to	B	228	ARG
A	107	PRO	to	B	232	ALA
A	110	ILE	to	B	233	ALA
A	110	ILE	to	B	232	ALA
A	110	ILE	to	B	229	ASN
A	111	PRO	to	B	252	ASP
A	111	PRO	to	B	253	GLY
A	111	PRO	to	B	232	ALA
A	114	HIS	to	B	233	ALA
A	114	HIS	to	B	234	GLY
A	114	HIS	to	B	232	ALA
A	114	HIS	to	B	172	ASP
A	114	HIS	to	B	170	TYR
A	120	LEU	to	B	233	ALA
A	121	GLU	to	B	235	SER
A	122	MET	to	B	233	ALA
A	122	MET	to	B	235	SER
A	122	MET	to	B	230	MET
A	123	TYR	to	B	219	PHE
A	123	TYR	to	B	230	MET
A	123	TYR	to	B	235	SER
A	123	TYR	to	B	218	HIS
A	124	LEU	to	B	226	THR
A	124	LEU	to	B	230	MET
A	124	LEU	to	B	229	ASN
A	124	LEU	to	B	219	PHE
A	124	LEU	to	B	233	ALA
A	124	LEU	to	B	221	ALA
A	125	HIS	to	B	219	PHE
A	125	HIS	to	B	226	THR
A	126	GLU	to	B	223	SER
A	126	GLU	to	B	222	THR
A	126	GLU	to	B	226	THR
A	126	GLU	to	B	225	GLU

Identification of Candidates

Gateway Cloning Technology (Invitrogen) was used to facilitate moving the mutated gene to other vectors. pGK702 contains the λcI OxyR repressor fusion. *oxyR* is flanked by the *att* sites for the Gateway system, effectively making this vector a Gateway expression vector. The mutated *oxy*R gene was moved into pDONR201 *via* the back reaction to generate a Gateway Entry Clone (Walhout et al. 2000). The entry clones were used to move the gene into several destination vectors: pLM1000 (Marino-Ramirez et al. 2004), pAZ299 and pGK751 (this study) *via* the LR reaction (see Table 2.3. for descriptions). The reactions were transformed into Mach1-T1^{R} cells. Candidate DNA was transformed into JH787 (Marino-Ramirez and Hu 2002) or AG1688 (Hu et al. 1993), streak purified and freezer stocks made.

Immunity Assay

To test for the repressor fusion's ability to oligomerize, cross-streak assays were done. Strains were struck out for single colonies and incubated for 16 hours at 37°C. Three individual colonies were challenged against lines of phage λKH4, λ*vir* and control phage λi^{21}c (Hu et al. 1990) on tryptone agar plates and incubated for 7 hours at 37°C. Those that were able to grow across the λKH54 line were called immune and those that were not able to grow were sensitive. All colonies died at the control λi^{21}c phage. Strains immune to λKH4 were given a score of one, while sensitive phenotypes received a score of zero. Each strain was tested three times with three individual colonies and the average calculated.

Table 2.3. Strains, Plasmids and Primers

AG1688	*ara*D139 Δ(*ara, leu*)7697, Δ*lacX*74, *gal*U, *gal*K, *hsd*R, *str*A F'128 *lac*I^q *lac*Z::Tn5
JH787	AG1688 Φ80 su3 supF
Mach1-T1R	F^- Δ*lac*X74 *hsd*R(r_k^- m_k^+) Δ*rec*1398 *end*A1 *ton*A
pAZ299	P_{7107}-cI-Gateway cassette
pDONR201	Gateway entry clone vector; contains *ccd*B, Cm^R, Km^R
pGK702	P_{7107}-cI-*att*B1-OxyR(2-305)-*att*B2, Amp^R
pGK751	P_{lacUV5}-cI-amber-Gateway cassette Cm^R, Amp^R
pLM1000	$P7_{107}$-cI-amber-Gateway cassette Cm^R, Amp^R
OxyR RD GW f	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAGATGGCAAGCCAGCAGGGCGAG-3'

Generation of λcI Regulatory Domain Repressor Fusions

Entry clones were generated of the 233 amino acids that comprise the regulatory domain of OxyR. Residues 80-305 of the regulatory domain of the full-length *oxy*R mutants were amplified with the primers OxyR RD GW-f and attB2 to attach in-frame *att*B sites. Using the PCR product, entry clones were generated in pDONR201 and the mutated genes subsequently moved to pLM1000, pAZ299 and pGK751 to generate λcI-regulatory domain repressor fusions.

Conservation Calculations

Ninety-seven OxyR orthologs were identified by using performing a BLAST search. We required the sequences have a cysteine at the equivalent *E. coli* positions of 199 and 208. A multiple sequence alignment was generated using the default settings on CLUSTAL-W. The different residues present at each position identified in QContacts were counted using the program QConAAtally.pl (see Appendix A) and the identity and conservation at the identified positions compared. Identical residues were scored for an exact match to the *E. coli K*-12 OxyR sequence. Conservation of residues were based on the following categories: Hydrophobic residues were VPWAILFM. Polar non-charged, GSTYCNQ. Positively charged positions were KRH and negatively charged residues were ED. A complete list of the orthologs is shown in Appendix B.

RESULTS

Identification of Interface Contact Residues

Residues in contact in both the reduced and oxidized forms of the OxyR were identified from the crystal structures as targets for alanine mutagenesis. Choi et al. described several residues that comprise the interfaces of the two forms of OxyR, including I110, H114, E121, L124, D172, H218, F219 and A233, but they do not present a comprehensive analysis of the interfaces. To provide such an analysis, we used the program QContacts to identify residues in Voronoi contact across the interfaces of the reduced and oxidized forms. The non-redundant contact profiles generated by QContacts are shown in Tables 2.2 and 2.3. The contact profile yields two types of information: the residues involved at the interface and the residues with which they interact on the opposite chain. When comparing these residues with the predicted residues from QContacts, we identify the specific residues described in the paper, as well as additional residues. We mapped these contacts on the reduced form of OxyR, as shown in Figure 2.1.

In the crystal structures solved by Choi et al. (Choi et al. 2001), the transition from the reduced to the oxidized form is mediated by a large rotation between the two subunits. To examine whether this involves formation of contacts between different sets of residues or changes in the contacts made by a common set of residues, we compared the contact profiles of the reduced and oxidized forms. Twenty-four residues were found in the contact profiles for both forms, while 18 are specific to only one of the two

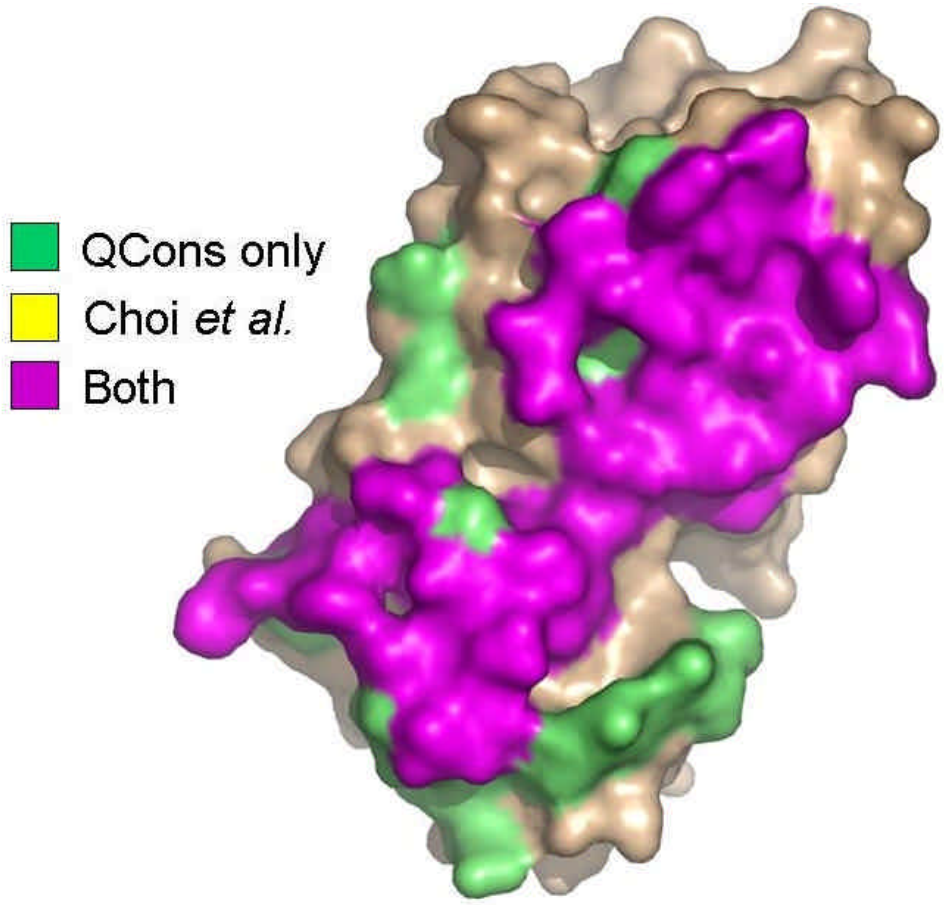


Figure 2.1. Comparison of Identified Structure Residues to QContacts. Residues detected by QContacts are green. Residues only described by Choi et al. (2000) are colored yellow. Residues colored purple are found in both studies.

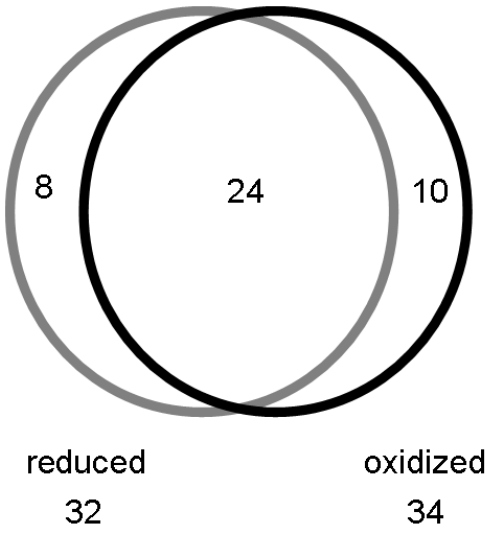
forms (Figure2.2A). However, when we examined the pair-wise contacts between the two profiles, only four are shared between the two forms (Figure2.2B). Additionally, though some atom-atom interactions are conserved, many are different. For example, in the reduced form, P103 Ca is making contacts with E225 Cß and Oe1, while in the oxidized form, Ca is not making any contacts (not shown).

Individual Residues' Contributions to Oligomerization

To systematically determine each residue's contribution to oligomerization, non-alanine or glycine residues were mutated to alanine. The mutant *oxyR* genes were subcloned into the λ repressor fusion vectors pLM1000, pAZ299 and pGK751. These vectors allow for increasing amounts of fusion protein to be expressed. Cross-streaks were performed to assay the fusions ability to oligomerize as described in the Materials and Methods. In all three vectors, full-length wild-type OxyR is immune. We sorted the immunity results of the full-length OxyR mutations into three classes: 1) Those that were not able to achieve immunity in any of the vectors, 2) those that were unable to oligomerize in pLM1000, but able to oligomerize in pAZ299 or pGK751 and 3) those that behaved as wild-type, being immune in all three vectors. These results are listed in Table 2.4.There are seven residues in class I, seventeen in class II and three in class III.

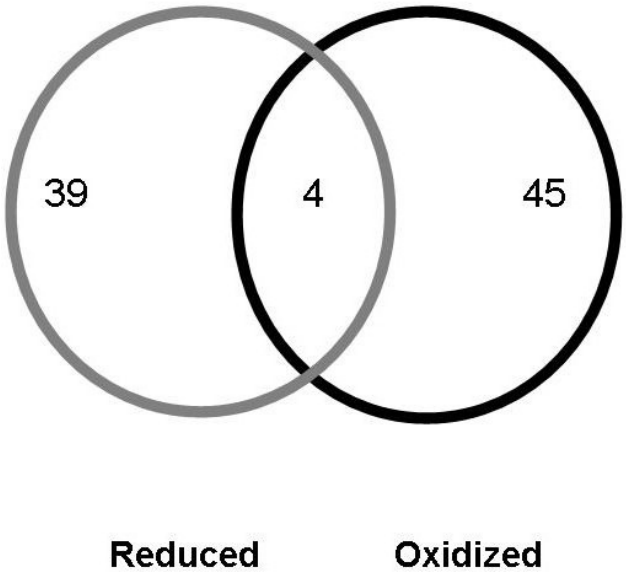
Though the crystal structure of OxyR is dimeric, the full length protein is tetrameric (Kullik et al. 1995a). Crystal structures of other LTTRs such as CbnR from *Ralstonia eutropha NH9* (Muraoka et al. 2003) and DntR from *Burkholderia* sp. strain DNT (Smirnova et al. 2004) show the tetramer as a dimer of dimers. To determine if the

A. Shared Residues



103	Pro	218	His
106	Leu	219	Phe
107	Pro	221	Ala
110	Ile	222	Thr
111	Pro	223	Ser
114	His	225	Glu
121	Glu	226	Thr
122	Met	229	Asn
123	Tyr	230	Met
124	Leu	233	Ala
125	His	235	Ser
126	Glu	252	Asp

B. Shared Contacts



103	Pro	To	225	Glu
107	Pro	To	229	Asn
123	Tyr	To	218	His
126	Glu	To	226	Thr

Figure 2.2. Comparison of the Contact Profiles of Reduced and Oxidized OxyR. Part A shows the residues found in both contact profiles. Part B shows the shared contacts found in both contact profiles of OxyR.

tetramer interactions were masking the dimeric interactions, we constructed λcI fusions to mutant and wild-type regulatory domains in all three vectors and assayed for immunity; the results are shown in Table 2.4. There are 20 class I, 8 class II and four class III residues. Comparing the full-length results to the regulatory results allowed for subcategorization. These are also listed in Table 2.4. The residues are illustrated in Figure 2.3.

Table 2.4. Immunity Result Classes

Position	Full-length	Regulatory Domain	Sub-class
E126	I	I	1A
R228	I	I	1A
E248	I	I	1A
H125	I	II	1B
H218	I	II	1B
M230	I	II	1B
S235	I	III	1C
Y104	II	I	2A
L106	II	I	2A
I110	II	I	2A
P111	II	I	2A
H114	II	I	2A
L120	II	I	2A
Q131	II	I	2A
T217	II	I	2A
S223	II	I	2A
E225	II	I	2A
D252	II	I	2A
Q128	II	II	2B
D172	II	II	2B
R220	II	II	2B
E121	II	III	2C
T226	II	III	2C
P107	III	I	3A
P103	III	I	3A
M122	III	I	3A
F219	III	I	3A
N229	III	I	3A
Y170	III	II	3B
T222	III	III	3C
H94	ND	II	
L97	ND	I	
L124	ND	I	

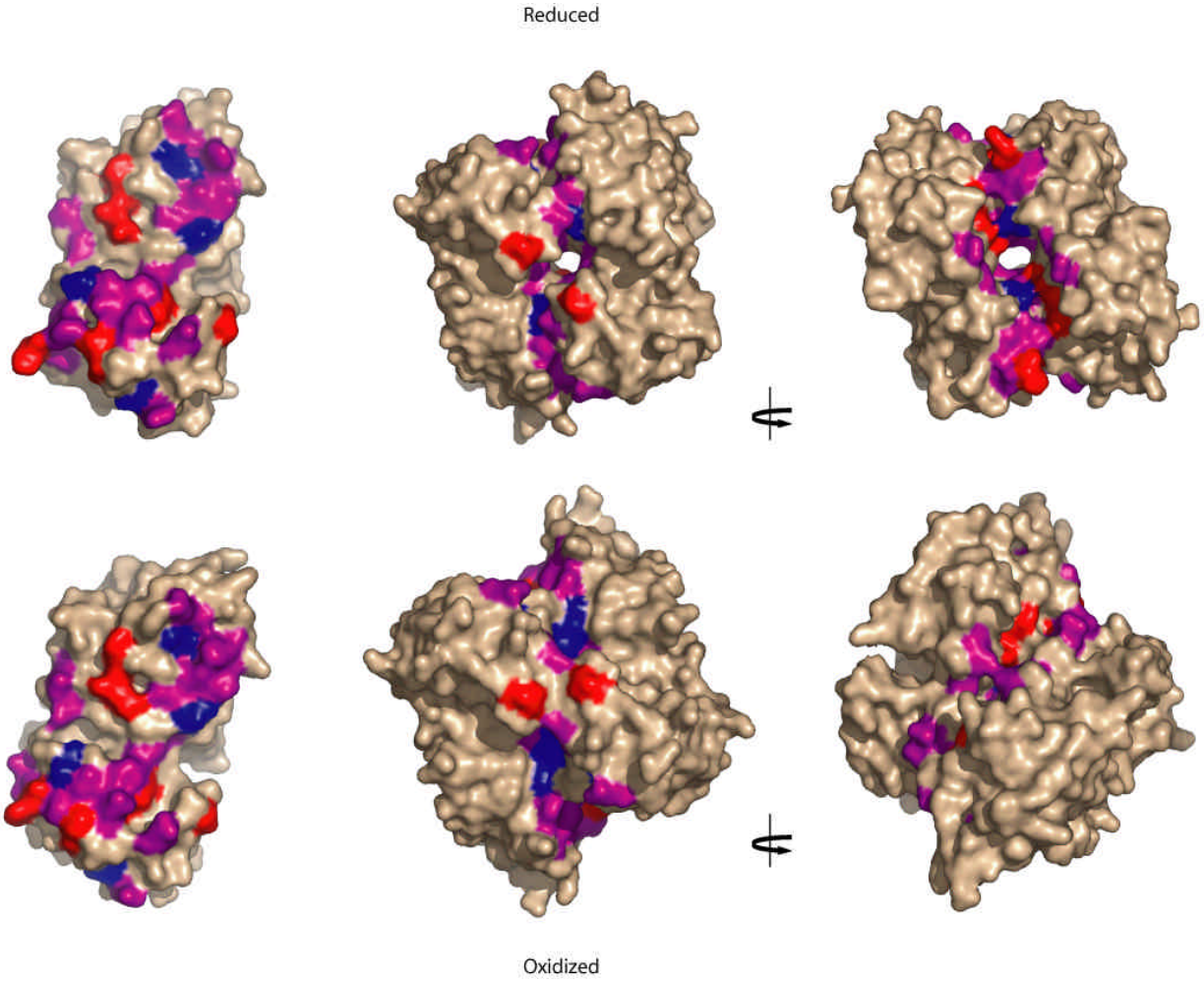


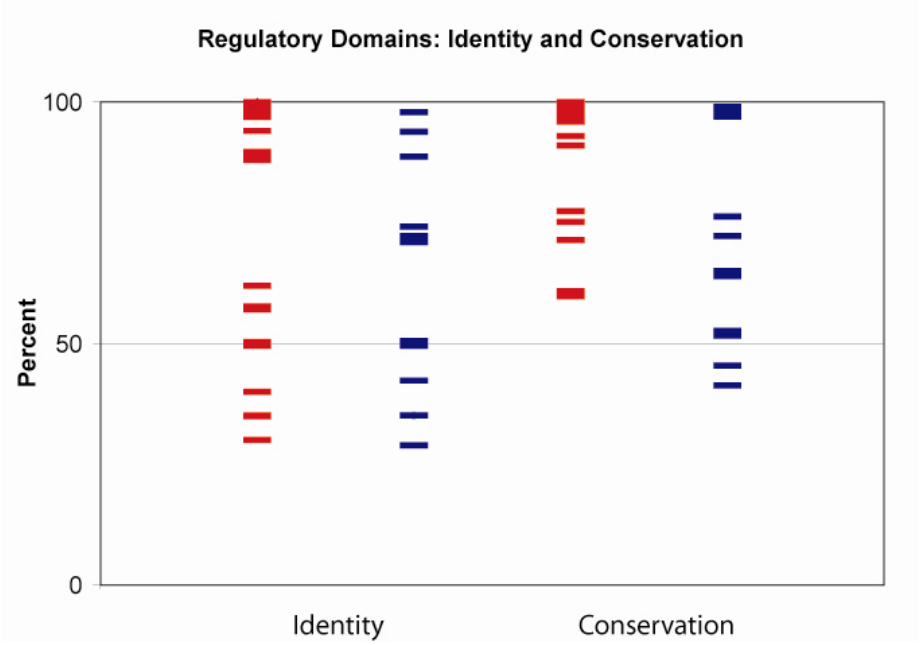
Figure 2.3. Mapped Classes of the Full-length Immunity Results. Red, Class I, purple Class II, blue Class III.

Conservation of Residues

To determine if there was a correlation between conservation of a residue amongst other OxyR orthologs and oligomerization, ninety-seven orthologs were identified (see Methods) and the number of different residues at each position calculated. Examination of the identity of the regulatory domain positions shows a broad range of numbers for both the residues that are important and not important. Figure 2.4A shows this range of identity from 20-100% (red bars). The same range of numbers is seen for residues that do not significantly contribute to oligomerization of OxyR (blue bars). Note that for simplicity, class II and class III residues were combined. The conservation of a residue was calculated and based on four categories of conservation: hydrophobic, polar uncharged, positively charged and negatively charged (see Methods). Considering conservation, the range, the range of percent conservation of the residues that contribute to the oligomerization largely overlaps with the residues that had no affect in oligomerization.

The same analysis was applied to the full-length residues. Considering identity only, the same range of percent identity found in the regulatory-domain results is found in both residues important and not important for oligomerization in the full-length (Figure 2.4B). Our conservation calculations are different than other's have used. Using the amino residues used by Keskin et al. and Guharoy and Chakrabarti, (Guharoy and Chakrabarti 2005; Keskin et al. 2005), we calculated conservation scores. The scores did not significantly alter our observed results (data not shown). Taken together, the residues important for oligomerization of OxyR are not more conserved than other interface

A. Regulatory-Domain: Identity and Conservation



B. Full-length: Identity and Conservation

Figure 2.4. Comparison of Identity and Conservation of Residues Based on Oligomerization Results. Each line represents a position that was designated important for oligomerization (red) or not important for oligomerization (blue).

residues.

DISCUSSION

In this study, we have combined *in silico* analysis and in vivo testing to address the oligomeric properties of OxyR. The residues targeted for alanine-scanning were generated from analysis using QContacts. There are residues unique to each contact profile, as well as shared residues. The shared residues, however, are involved in different contacts across the subunit, indicative of the large rotation between the subunits in the reduced versus oxidized form. One example is seen with described interactions by Choi et al. In the reduced form, A233 is interacting with I110 and L124. In the oxidized form, I110 and L124 are interacting with F219 (Choi et al. 2001). These interactions are detected by QContacts.

Utilization of repressor fusions enabled a systematic study of an individual residue's contribution to oligomerization. Testing the contribution of each residue in both a full-length and regulatory domain repressor fusion allowed us to identify residues that behave differently in each state. The regulatory domain removes the tetrameric interactions from the system, identifying residues that contribute to oligomerization, though not as much as residues identified by the full-length experiments (see classes 2A, 3A, 2B and 3B) and essentially validate the residues generated from QContacts.

Seven residues were identified as being important for the oligomerization of OxyR. We calculated the conservation at these positions in OxyR orthologs and did not find a correlation between conservation and oligomerization. This is uncharacteristic of

other hot spots found in literature. If we consider the characteristics described in Chapter I, the hot spots identified in this study are not typical hot spots. While they appear to largely contribute to the oligomerization, they are not interacting with hot spots on the opposite subunit. Further, while they are polar in nature by Hu et al. standards, none are tryptophan, tyrosine and only one is arginine; residues for which hot spots are especially enriched (Bogan and Thorn 1998; Hu et al. 2000).

Keskin et al. (2005), as did Bogan and Thorn (1998), found that the hot spots tended to be clustered together. Keskin et al. termed these hot regions. Further, these regions tended to be protected from bulk solvent and located toward the center of the interface. We find that this is quite the opposite of our relevant OxyR positions. As seen in Figure 2.3, the hot spots are distributed over the surface of the interface. Finally, the hot spots in OxyR are not especially conserved. This, by far, was the most surprising result of this study, as several studies have emphasized the observation that the hot spots are conserved (Tsai et al. 1996; Hu et al. 2000).

The large rotation between the subunits observed in the transition from the reduced to oxidized form may provide an explanation for the uncharacteristic properties of the hot spots observed in this study. With such a large rotation, having all hot spots toward the center of the molecule may inhibit the molecule's ability to undergo this conformationl change. Closer examination of the hot spots' contacts in each crystal form illustrate this point. As shown in Table 2.5, in the reduced form, M230 makes interactions with M122, Y123 and L124. In the reduced form, these interactions are

replaced with interacting residues L106, P107, I110 and P111. Additional examples are found in other hot spots' contacts.

Table 2.5. Interactions of M230 in the Reduced and Oxidized Forms

REDUCED	
M230	M122
	Y123
	L124

OXIDIZED	
M230	L106
	P107
	I110
	P111

Five of the seven hot spots are involved in contacts that are shared between the two forms, while E248 is found only in the oxidized contact profile and R228 is found only in the reduced contact profile. While we have a working assumption the cell is under a reduced state, the assay is in vivo and the cells are being exposed to phage. This exposure may influence the conformation of the protein, allowing for residues in the oxidized profile to be detected.

Subclasses 1B and 1C are interesting in so much that they are sensitive as full-lengths and immune in the regulatory domain. While we do not understand what is happening, one possibility may lie in the oxidation state of the protein, as this has not been addressed. We attempted to do the repressor assays under a constitutively oxidized state; however, control assays did not behave as expected and thus, the results were not

interpretable (data not shown). While we have a working assumption that the cell is in a reductive state, we can not be sure. By extension, we do not know the conformation of OxyR. Thus, subclasses 1B and 1C may be explained by a change in oxidation.

Subclass 3C is uniquely interesting. It is immune in both the full-length and regulatory domain. We examined the location of the position, T222, and found that it was located toward the middle of the interface. Examination of the interacting atoms shows all atoms in T222 interacting with residues in V97, L106, L124, H125 and E126 in the oxidized form and E126 in the reduced form. H125 and E126 are both hot spot residues.

CHAPTER III

THE OLIGOMERIZATION PROPERTIES OF CynR

OVERVIEW

Deletion analysis and alanine-scanning based on a homology-based interaction model were used to identify determinants of oligomerization in the transcriptional regulator CynR, a member of the LysR-Type Transcriptional Regulator (LTTR) family. Deletion analysis confirmed that the putative regulatory domain of CynR was essential for driving the oligomerization of λ χI repressor-CynR fusion proteins. The interaction surface of a different LTTR, OxyR, was mapped onto a multiple sequence alignment of the LTTR family. This mapping identified putative contacts in the CynR regulatory domain dimer interface, which were targeted for alanine scanning mutagenesis.

INTRODUCTION

Most LTTRS are dimers of dimers, and oligomerization is essential for their function as transcriptional activators. In the full-length structure of CbnR, the first LTTR for which a full-length structure was determined, a coiled-coil links the DNA-binding domain to the regulatory domain within one monomer (Muraoka et al. 2003). There are two possible interactions among the monomeric subunits in dimer formation: one at the regulator domain of subunits B and P; the other at the a-helical linker of B and A.

Many organisms encode multiple LTTRs in their genomes. There are 47 in *E. coli* and 123 in *P. aeruginosa*. Because the oligomerization of the LTTRs shows high

specificity (Chapter IV), the oligomerization interfaces must have diverged within a common structural framework to achieve specificity while maintaining stability. To understand the evolution of specificity in protein interactions, we are analyzing the basis of regulatory domain dimerization in LTTRs. Previously (Chapter II), we used the program QContacts and alanine-scanning mutagenesis to identify residues that contribute to the oligomerization of OxyR, identifying seven hot spot residues.

CynR is the transcriptional regulator of the *cyn* operon, which encodes genes that allow cyanate to be used as a sole-source of nitrogen. The operon consists of *cynT, cynS, cynX*, which encode a carbonic anhydrase, a cyanase and a protein of unknown function, respectively (Sung and Fuchs 1988). CynR binds to DNA in vitro in the presence or absence of cyanate. CynR binding induced bending of the DNA under both conditions, but the amount of DNA bending was decreased in the presence of cyanate (Lamblin and Fuchs 1994). Unlike OxyR, a structure for the CynR regulatory domain was not available when this study was begun.

Here, we describe studies on the oligomerization determinants of CynR, using the domain structure of CbnR and analysis of the interaction surface of the OxyR regulatory domain dimer to guide deletions and alanine scanning mutagenesis. The results reveal interesting similarities and differences between OxyR and CynR. During the course of this study, the structure of the regulatory domain of CynR without the inducer was deposited in the PDB (2HXR), allowing us to evaluate the performance of the homology-driven targeting of sites for mutations.

METHODS

Construction of Deletion Mutations

Relevant strains, primers and plasmids are described in Table 3.1. To determine the equivalent positions of CbnR in CynR, a multiple sequence alignment of COG0583 (Tatusov et al. 2000) and CbnR from *R. eutrophea* (Muraoka et al. 2003) was generated using CLUSTAL W (1.82). The regions used to determine the constructs are based on those described in the crystal structure paper of CbnR. Oligos were designed to attach in-frame *Sal*I and *Bam*HI sites and ordered from IDT (Iowa). These oligos were used to amplify the appropriate fragments out of full-length CynR (pGK304) and checked for the appropriate size on a 1.8% agarose gel, and digested with *Sal*I and *Bam*HI. The inserts were cleaned using a Qiagen PCR Clean-up Kit and ligated into digested pJH391 digested with *Sal*I and *Bam*HI, generating λcI repressor fusions. Ligations were transformed into AG1688 and transformants were recovered and sequenced (LPGT, TAMU).

Immunity Assay

Cross-streak assays were used to test the ability of the λcI fusions to oligomerize. In general, cells were streaked out for single colonies and incubated for 16 hours at 37°C. Individual colonies were challenged against lines of phage λKH4, λ*vir* and control phage $\lambda i^{21}c$ (Hu et al. 1990) on tryptone agar plates and incubated for 7 hours at 37°C. Those that were able to grow across the λKH54 line were called immune and those that were not able to grow were sensitive. All colonies died at the control $\lambda^{21}c$

phage. For the deletion mutation cross-streaks, stains were simply called immune or sensitive, based on their phenotype to λKH54. For other described cross-streaks, strains immune to λKH4 were given a score of one, while sensitive phenotypes received a score of zero. Each strain was tested three times with three individual colonies. Immune colonies were given a score of 1, with sensitive colonies given a score of 0. The average of all experiments was calculated to give an immunity score.

Dimerization and Tetramerization Test

To distinguish between cI repressor fusions of dimer and tetramers, the clones were moved into JH607 and XZ980 *via* M-13 transduction. JH607 and XZ980 are strains that are used to distinguish cooperative oligomerization and have been previously described (Zeng and Hu 1997). Briefly, JH607 contains the construct $\lambda 112O_sP_s$, which contains a synthetic promoter that drives expression of *cat* and *lacZ.* A weak lambda operator overlaps the promoter and a strong operator is upstream of the weaker promoter. Strong repression is only detected in higher order oligomerization states. XZ980 contains the same weak promoter. The strong upstream operator was replaced with a λ434 operator. ß-galactosidase assays were done according to Miller (Miller 1972).

Generation of Contact Profiles

In chapter II, the use of QContacts to generate a contact profile of the two forms of OxyR utilizing the available crystal structures of the regulatory domain was

described. The combined contact profiles generated from the OxyR study were used as a template to generate predicted CynR contact profiles. The multiple sequence alignment described above was used. A PERL script called QConalign.pl (Appendix A) was written to parse the residues at equivalent positions in OxyR out of CynR using the multiple sequence alignment, generating a predicted contact profile for CynR.

Alanine Scanning Mutagenesis

Using the contact profiles as a guide, the identified non-alanine and glycine residues were mutated to alanine. PrimerX (http://www.bioinformatics.org/primerx) was used to generate the primer pairs and primers were ordered from IDT (Iowa). Pfu Turbo (Stratagene) or Pfx Platinum (Invitrogen) was used for the polymerase. pGK343 was the DNA template. Reactions were treated with 2U of *Dpn*I for 2-6 hours and transformed into Mach1-T1^R, Top1-T1^R (Invitrogen) or XL-1Blue Supercompetent (Stratagene) cells. Transformants were picked and streak purified, cultured and plasmid recovered and sequenced (LPGT, TAMU).

Generation of λcI Repressor Fusions with Different Expression Levels

For each mutant, we generated a Gateway Entry Clone (Walhout et al. 2000). pGK343 is a λcI repressor fusion of full-length CynR with Gateway attachment sites flanking the full-length CynR. The mutated *cynR* gene was moved into pDONR201 *via* the back reaction to generate the entry clone, transformed into Mach1-T1^R cells and selected on LB plates containing kanamycin. Candidates were screened for the loss of

ampicillin resistance by streak purification on LB kanamycin and LB ampicllin plates. The entry clones were used to move the gene into several destination vectors: pLM1000 (Marino-Ramirez et al. 2004), pAZ299 and pGK751 *via* the LR reaction. The reactions were transformed into Mach1-T1^{R} cells. Candidates were screened for loss of grown on LB kanamycin, LB chlorepnenicol and ability to grow on LB ampicillin. All constructs were subjected to restriction mapping to ensure the appropriate clone was generated. Depending on the fusion, the plasmid was transformed into either JH787 (Marino-Ramirez and Hu 2002) or AG1688 (Hu et al. 1993).

Generation of λcI Regulatory Domain Fusions

Entry clones were generated of the mutated regulatory domains. Residues 89-299 of the regulatory domain of the full-length CynR mutants were amplified with the CynR RD GW-f and CynR RD GW-r or *att*B2 in-frame *att*B sites. Using the PCR product, entry clones were generated in pDONR201. To generate λcI regulatory domain fusions, most of the mutated genes were subsequently moved to pLM1000, pAZ299 and pGK751.

Growth Curves

2mL overnights of strains deleted for *cynR, cynT* or *cynS* and the parental BW25113 were grown in LB with aeration at 37°C. Cultures were typically diluted 1:200 into 25mL of pre-warmed Minimal A + 0.2% glucose + AAA-Arg. Cells were grown in a 37°C shaking water bath until an OD_{600} of 0.2 was reached. Freshly made

KCON was added to 1mM or 5mM concentrations and growth monitored by removing .5mL samples and measuring the OD_{600}.

Table 3.1. Relevant Strains, Primers and Plasmids

pGK304	P_{7107}-cI-amber-CynR
pGK343	P_{7107}-cI-amber-att-CynR-att
pJH391	P_{lacUV5}-cI vector
pLM1000	P_{7107}-cI-amber-Gateway® cassette
pGK751	P_{lacUV5}-cI-amber-Gateway® cassette
pAZ299	P_{7107}-cI-Gateway® cassette
AG1688	*araD*139 Δ(*ara, leu*)7697, Δ*lacX*74*, galU, galK, hsdR*, *strA* F'128 lacI^q *lacZ::Tn5*
JH787	AG1688 Φ80 su3 supF
JH607	AG1688 λ112O_sP_s
XZ980	AG1688 λXZ970
Mach1-T1	F$^-$ Δ*lac*X74 *hsd*R(r_k^- m_k^+) Δ*rec*1398 *end*A1 *ton*A
CynR RD-GW-f	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTATCTGA CGCGAGGATCGCTG
CynR RD GW-r	5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACC GTGATTCATTTCCGCCAA-3'
BW25113	*D(araD-araB)567, DlacZ4787(::rrnB-3), lambda$^-$, rph-1, D(rhaD-rhaB)568, hsdR514*
JW0311	BW25113 cynS781(del)::kan
JW0330	BW25113 ΔcynT780::kan
JW5894	BW25113 ΔcynR::kan

RESULTS

Regions Sufficient for Oligomerization

The crystal structure of CbnR showed interactions between all four subunits. If named A, B, P and Q, interactions between B and P at the C-terminal region were observed, with interactions between subunits A and B at the helical linker. To determine which portions of the CynR protein were necessary and sufficient for oligomerization, deletion mutations of CynR were fused to the DNA binding domain of λcI to construct λ repressor fusions. The deletion mutations constructed are shown in Figure 3.1. The repressor fusions were challenged against phage λKH54 and immunity assayed.

Determination of Oligomerization State

In the CbnR structure, the tetramer consists of a dimer of dimers. To determine whether or not the immune fusions were forming dimers or higher order oligomers, the clones were moved into *lacZ* reporter strains that can distinguish between monomers and higher order oligomers (Zeng and Hu 1997) and the ability to repress the reporter was assayed using ß-galactosidase assays; results are shown in Table 3.2. pZ150 is the empty vector. pKH101 is the N-terminal DNA binding domain of λcI repressor. The constructs containing only the full regulatory domain of CynR (Table 3.2, constructs F and G) showed repression levels similar to those of the dimeric control (pJH370). Only the full-length (Table 3.2, construct E) showed repression levels similar to that of the tetrameric control (pJH157). Taken together with the cross-streak assays, these results indicate that the full regulatory domain is sufficient for dimerization while the whole protein is necessary for tetramerization.

CbnR	1-57	58-87	88-165	166-259	260-294	
CynR	1-57	58-88	89-165	166-257	258-299	
	DNA Binding Domain	**Helical Linker**	**Regulatory Domain**			**Immunity**
A						S
B						S
C						S
D						I
E						I
F						I
G						I
H						S
I						S
J						S

Figure 3.1. Constructed Deletion Mutations. The amino acid regions in CbnR are described in the top row. In the second row, the regions in CynR that were constructed are described. The regions included in each construct (A-J) are shaded under the described included amino acid region.

Table 3.2. Assay for Higher Order Oligomers

Construct	**JH607**	**XZ980**
pZ150	100% ± 0	100% ± 0
pKH101	34 ± 7	116 ± 44
pJH157 (tetramer)	4 ± 2	78 ± 24
pJH370 (dimer)	45 ± 21	45 ± 16
D	80 ± 44	115 ± 48
E	7 ± 5	69 ± 4
F	38 ± 17	85 ± 5
G	34 ± 10	81 ± 24

Prediction of the CynR Interface

Because of the overall similarity in the structures of the regulatory domains of LTTRs, we hypothesized that positions important for oligomerization in one LTTR may be important in the oligomerization of another LTTR. In a previous study of OxyR oligomerization, a set of residues that contribute to oligomerization through the use of alanine-scanning was identified. Using the multiple sequence alignment described in the Methods and the program QConAlign.pl (Appendix A), the equivalent positions in the contact profiles identified in the OxyR study in CynR were determined, generating predicted contact profiles for CynR, as shown in Tables 3.3 and 3.4.

The residues at the positions of the contact profiles were compared between OxyR and CynR. Of the 42 residues in the OxyR contact profile, six positions were identical in both proteins: Y104, L124, E126, A221, S223 and S235. Three positions, 251, 252 and 253, were not present in the CynR contact profile.

Determination of Real CynR Interface

During the course of this study, the structure of the regulatory domain of CynR without the inducer was deposited in the PDB (2HXR). The unreleased structure of the regulatory domain of CynR with sodium azide bound was graciously provided to us by Savchenko and colleagues, allowing for the actual contacts of the CynR structure to be calculated using QContacts, generating a real contact profile for each form of the CynR structures. These contacts are shown in Tables 3.5 and 3.6.

Table 3.3 Predicted Contacts in CynR Based on Reduced OxyR

Chain	Residue #	Name	Contacts	Chain	Residue #	Name
A	103	S	to	B	225	S
A	103	S	to	B	229	E
A	106	I	to	B	229	E
A	106	I	to	B	225	S
A	106	I	to	B	226	A
A	107	G	to	B	252	-
A	107	G	to	B	229	E
A	107	G	to	B	228	L
A	107	G	to	B	232	R
A	110	M	to	B	233	R
A	110	M	to	B	232	R
A	110	M	to	B	229	E
A	111	A	to	B	252	-
A	111	A	to	B	253	-
A	111	A	to	B	232	R
A	114	Y	to	B	233	R
A	114	Y	to	B	234	T
A	114	Y	to	B	232	R
A	114	Y	to	B	172	H
A	114	Y	to	B	170	A
A	120	I	to	B	233	R
A	122	L	to	B	233	R
A	122	L	to	B	235	S
A	122	L	to	B	230	L
A	122	L	to	B	218	V
A	123	Q	to	B	219	I
A	123	Q	to	B	230	L
A	123	Q	to	B	235	S
A	123	Q	to	B	218	V
A	124	L	to	B	226	A
A	124	L	to	B	230	L
A	124	L	to	B	229	E
A	124	L	to	B	219	I
A	124	L	to	B	233	R
A	124	L	to	B	221	A
A	125	Q	to	B	219	I
A	125	Q	to	B	226	A
A	126	E	to	B	223	S
A	126	E	to	B	222	N
A	126	E	to	B	226	A
A	126	E	to	B	225	S

Table 3.4. Predicted Contacts in CynR Based on Oxidized OxyR

Chain	Residue #	Name	Contacts	Chain	Residue #	Name
A	97	V	to	B	222	N
A	102	T	to	B	226	A
A	103	S	to	B	226	A
A	103	S	to	B	223	S
A	103	S	to	B	229	E
A	103	S	to	B	226	A
A	104	Y	to	B	229	E
A	104	Y	to	B	225	S
A	106	I	to	B	230	L
A	106	I	to	B	222	N
A	106	I	to	B	219	I
A	106	I	to	B	226	A
A	107	G	to	B	229	E
A	107	G	to	B	230	L
A	107	G	to	B	233	R
A	107	G	to	B	226	A
A	110	M	to	B	235	S
A	110	M	to	B	230	L
A	110	M	to	B	219	I
A	111	A	to	B	233	R
A	111	A	to	B	230	L
A	111	A	to	B	235	S
A	114	Y	to	B	235	S
A	114	Y	to	B	219	I
A	121	T	to	B	235	S
A	121	T	to	B	218	V
A	122	L	to	B	218	V
A	123	Q	to	B	218	V
A	123	Q	to	B	217	V
A	123	Q	to	B	220	E
A	124	L	to	B	221	A
A	124	L	to	B	222	N
A	124	L	to	B	218	V
A	124	L	to	B	220	E
A	125	Q	to	B	221	A
A	125	Q	to	B	222	N
A	126	E	to	B	223	S
A	126	E	to	B	226	A
A	126	E	to	B	128	S
A	131	K	to	B	131	K
A	225	S	to	B	225	S
A	247	Q	to	B	252	-
A	248	H	to	B	252	-
A	251	-	to	B	251	-

Table 3.5. Contact Profile of Uninduced CynR

Res. Num	Residue	Chain	Contacts	Res. Num	Residue	Chain
102	THR	A	to	225	SER	B
102	THR	A	to	229	GLU	B
102	THR	A	to	226	ALA	B
102	THR	A	to	223	SER	B
103	SER	A	to	225	SER	B
106	ILE	A	to	230	LEU	B
106	ILE	A	to	229	GLU	B
106	ILE	A	to	226	ALA	B
107	GLY	A	to	229	GLU	B
107	GLY	A	to	230	LEU	B
107	GLY	A	to	226	ALA	B
107	GLY	A	to	233	ARG	B
108	PRO	A	to	233	ARG	B
108	PRO	A	to	229	GLU	B
111	ALA	A	to	230	LEU	B
111	ALA	A	to	234	THR	B
111	ALA	A	to	219	ILE	B
111	ALA	A	to	233	ARG	B
114	TYR	A	to	236	LEU	B
114	TYR	A	to	234	THR	B
114	TYR	A	to	219	ILE	B
114	TYR	A	to	218	VAL	B
114	TYR	A	to	190	LYS	B
115	ALA	A	to	234	THR	B
118	PRO	A	to	218	VAL	B
118	PRO	A	to	190	LYS	B
120	ILE	A	to	218	VAL	B
120	ILE	A	to	219	ILE	B
121	THR	A	to	218	VAL	B
121	THR	A	to	217	VAL	B
121	THR	A	to	220	GLU	B
121	THR	A	to	219	ILE	B
122	LEU	A	to	219	ILE	B
122	LEU	A	to	218	VAL	B
122	LEU	A	to	220	GLU	B
122	LEU	A	to	230	LEU	B
123	GLN	A	to	220	GLU	B
123	GLN	A	to	219	ILE	B
123	GLN	A	to	197	GLU	B
123	GLN	A	to	195	SER	B
124	LEU	A	to	221	ALA	B

Table 3.5. Continued

Res. Num	Residue	Chain	Contacts	Res. Num	Residue	Chain
124	LEU	A	to	220	GLU	B
124	LEU	A	to	219	ILE	B
124	LEU	A	to	222	ASN	B
124	LEU	A	to	230	LEU	B
124	LEU	A	to	226	ALA	B
125	GLN	A	to	222	ASN	B
125	GLN	A	to	195	SER	B
125	GLN	A	to	197	GLU	B
126	GLU	A	to	222	ASN	B
126	GLU	A	to	223	SER	B
223	SER	A	to	124	LEU	B
225	SER	A	to	225	SER	B
229	GLU	A	to	247	GLN	B
229	GLU	A	to	104	TYR	B
229	GLU	A	to	103	SER	B
232	ARG	A	to	248	HIS	B
233	ARG	A	to	247	GLN	B
248	HIS	A	to	229	GLU	B

Table 3.6. Contact Profile of Induced CynR

Res. Num	Residue	Chain	Contacts	Res. Num	Residue	Chain
102	THR	A	to	223	SER	C
102	THR	A	to	225	SER	C
102	THR	A	to	226	ALA	C
102	THR	A	to	229	GLU	C
103	SER	A	to	225	SER	C
103	SER	A	to	229	GLU	C
104	TYR	A	to	229	GLU	C
106	ILE	A	to	226	ALA	C
106	ILE	A	to	230	LEU	C
107	GLY	A	to	230	LEU	C
107	GLY	A	to	229	GLU	C
107	GLY	A	to	226	ALA	C
107	GLY	A	to	233	ARG	C
108	PRO	A	to	229	GLU	C
108	PRO	A	to	233	ARG	C
111	ALA	A	to	230	LEU	C
111	ALA	A	to	234	THR	C
111	ALA	A	to	233	ARG	C
114	TYR	A	to	230	LEU	C
114	TYR	A	to	219	ILE	C
114	TYR	A	to	234	THR	C
114	TYR	A	to	236	LEU	C
114	TYR	A	to	218	VAL	C
114	TYR	A	to	190	LYS	C
115	ALA	A	to	234	THR	C
115	ALA	A	to	233	ARG	C
118	PRO	A	to	190	LYS	C
118	PRO	A	to	218	VAL	C
120	ILE	A	to	218	VAL	C
120	ILE	A	to	219	ILE	C
121	THR	A	to	218	VAL	C
121	THR	A	to	217	VAL	C
121	THR	A	to	216	GLN	C
122	LEU	A	to	219	ILE	C
122	LEU	A	to	218	VAL	C
122	LEU	A	to	230	LEU	C
122	LEU	A	to	220	GLU	C
123	GLN	A	to	220	GLU	C

Table 3.6. Continued

Res. Num	Residue	Chain	Contacts	Res. Num	Residue	Chain
123	GLN	A	to	221	ALA	C
123	GLN	A	to	195	SER	C
124	LEU	A	to	220	GLU	C
124	LEU	A	to	221	ALA	C
124	LEU	A	to	230	LEU	C
124	LEU	A	to	219	ILE	C
124	LEU	A	to	226	ALA	C
124	LEU	A	to	223	SER	C
124	LEU	A	to	222	ASN	C
125	GLN	A	to	222	ASN	C
125	GLN	A	to	197	GLU	C
125	GLN	A	to	221	ALA	C
125	GLN	A	to	195	SER	C
126	GLU	A	to	222	ASN	C
126	GLU	A	to	223	SER	C
190	LYS	A	to	119	SER	C
190	LYS	A	to	118	PRO	C
197	GLU	A	to	123	GLN	C
225	SER	A	to	225	SER	C
229	GLU	A	to	247	GLN	C
232	ARG	A	to	248	HIS	C
233	ARG	A	to	247	GLN	C
248	HIS	A	to	248	HIS	C

The two contact profiles based on the CynR structures were compared for similarities and differences. We first looked at residue similarities and found that all the residues contributing to the interface of the uninduced contribute to the interface of the induced. Only two residues were found in the induced structure that were not in the uninduced structure: 119 and 216.

Next, the individual contacts contributing to the interface were compared. Overall, the uninduced structure had 59 contacts, while the induced structure had 60 contacts. Compared together, they shared 53 contacts. The uninduced had six unique while induced had seven (Figure 3.2, Top). However, examination of the actual atomic level interactions reveals differences between the atoms in the contacts. For example, in both crystal forms, L106 interacts with A226. In the uninduced structure, Cß of the alanine interacts with Cd1 of the ILE. In the induced structure, Cß is now interacting with Cγ2. An additional contact of the Ca of the alanine is also interacting with Cγ2 (Figure 3.2, Bottom).

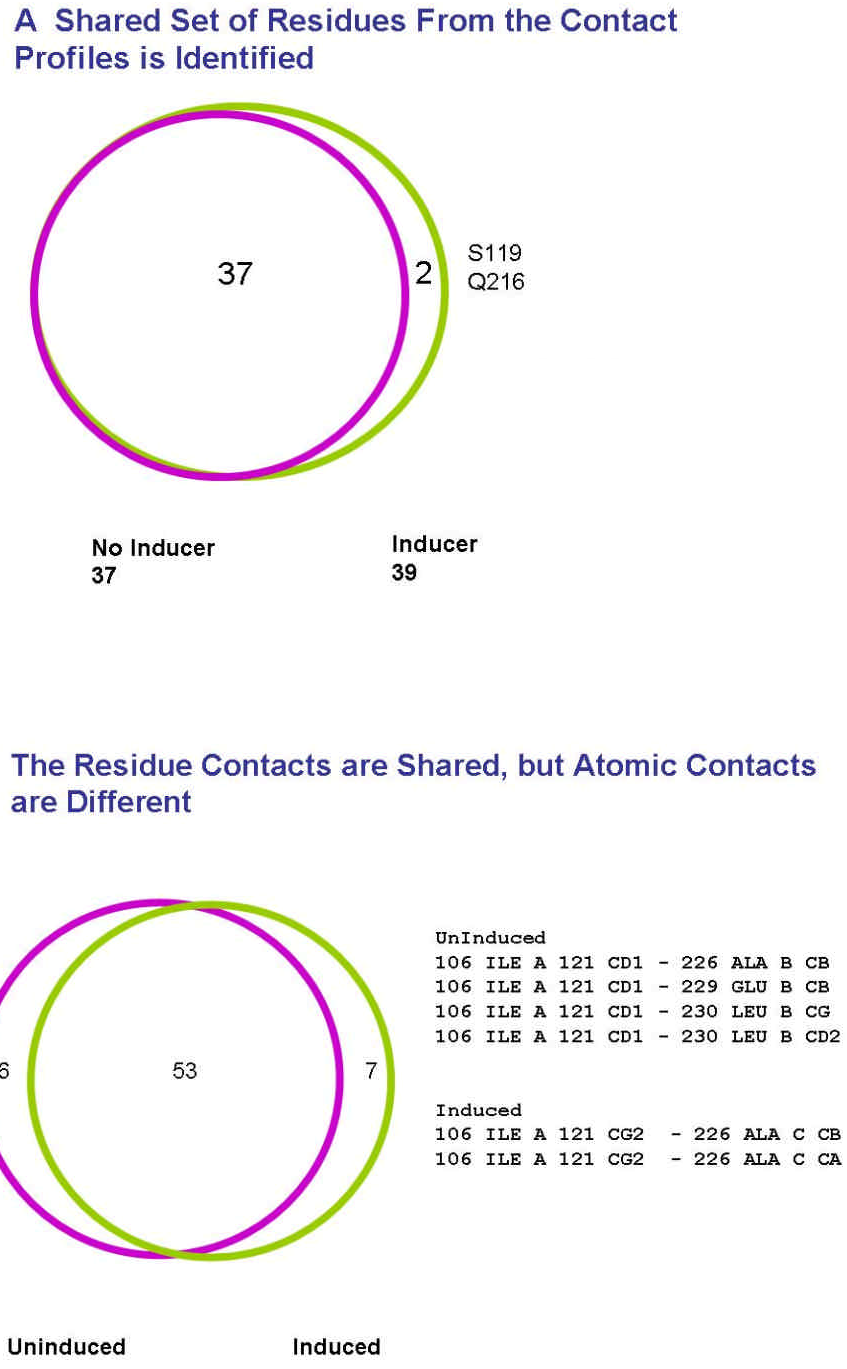


Figure 3.2. Comparison of the Contact Profiles of CynR. The top portion shows the residues found in both contact profiles of CynR based on the crystal structures. The bottom shows the contacts that are shared between the contact profiles. The atomic contacts, however, are different.

Comparisons of All Contact Profiles

The list of residues from the predicted contact profiles were combined and compared to the real contact profiles of the uninduced and induced forms of CynR. Residues shared by the real and predicted contact profiles are shown in Table 3.7. Of the 48 residues amongst all four contact profiles, 21 residues were found in all contact profiles. Residues unique to the contact profiles based on OxyR were: R94, V97, M110, S128, K131, A170, H172 and L228. Residues unique to the contact profiles based on the CynR crystal structures were: P108, A115, P118, S119, K190, S195, E197, Q216 and L236

To illustrate why some residues predicted from the OxyR based contact profile were not detected in the contract profiles based on the CynR structures, the residues found in all contact profiles were mapped onto the surface of the structures of CynR (Figure 3.3). Further, the residues found in only the contact profiles detected form the CynR contact profiles and the predicted contact profiles are also noted on the structure of CynR. The shared residues map to the interface of CynR (purple color).

Individual Residues' Contributions to Oligomerization

To determine a residue's contribution to oligomerization, non-alanine or glycine residues were mutated to alanine. The mutant *cynR* genes were moved to pLM1000 and

Table 3.7. Residues from the Four Contact Profiles of CynR

Predicted from OxyR Reduced	Predicted from OxyR Oxidized	CynR Uninduced	CynR Induced
R94	V97		
	T102	T102	T102
S103	S103	S103	S103
	Y104	Y104	Y104
I106	I106	I106	I106
G107	G107	G107	G017
M110	M110	P108	P108
A111	A111	A111	A111
Y114	Y114	Y114	Y114
		A115	A115
		P118	P118
			S119
I120		I120	I120
	T121	T121	T121
L122	L122	L122	L122
Q123	Q123	Q123	Q123
L124	L124	L124	L124
Q125	Q125	Q125	Q125
E126	E126	E126	E126
A170	S128	K190	K190
H172	K131	S195	S195
		E197	E197
			Q216
	V217	V217	V217
V218	V218	V218	V218
I219	I219	I219	I219
	E220	E220	E220
A221	A221	A221	A221
N222	N222	N222	N222
S223	S223	S223	S223
S225	S225	S225	S225
A226	A226	A226	A226
L228			
E229	E229	E229	E229
L230	L230	L230	L230
R232	R232	R232	R232
R233	R233	R233	R233
T234		T234	T234
S235	S235		
		L236	L236
	Q247	Q247	Q247
	H248	H248	H248

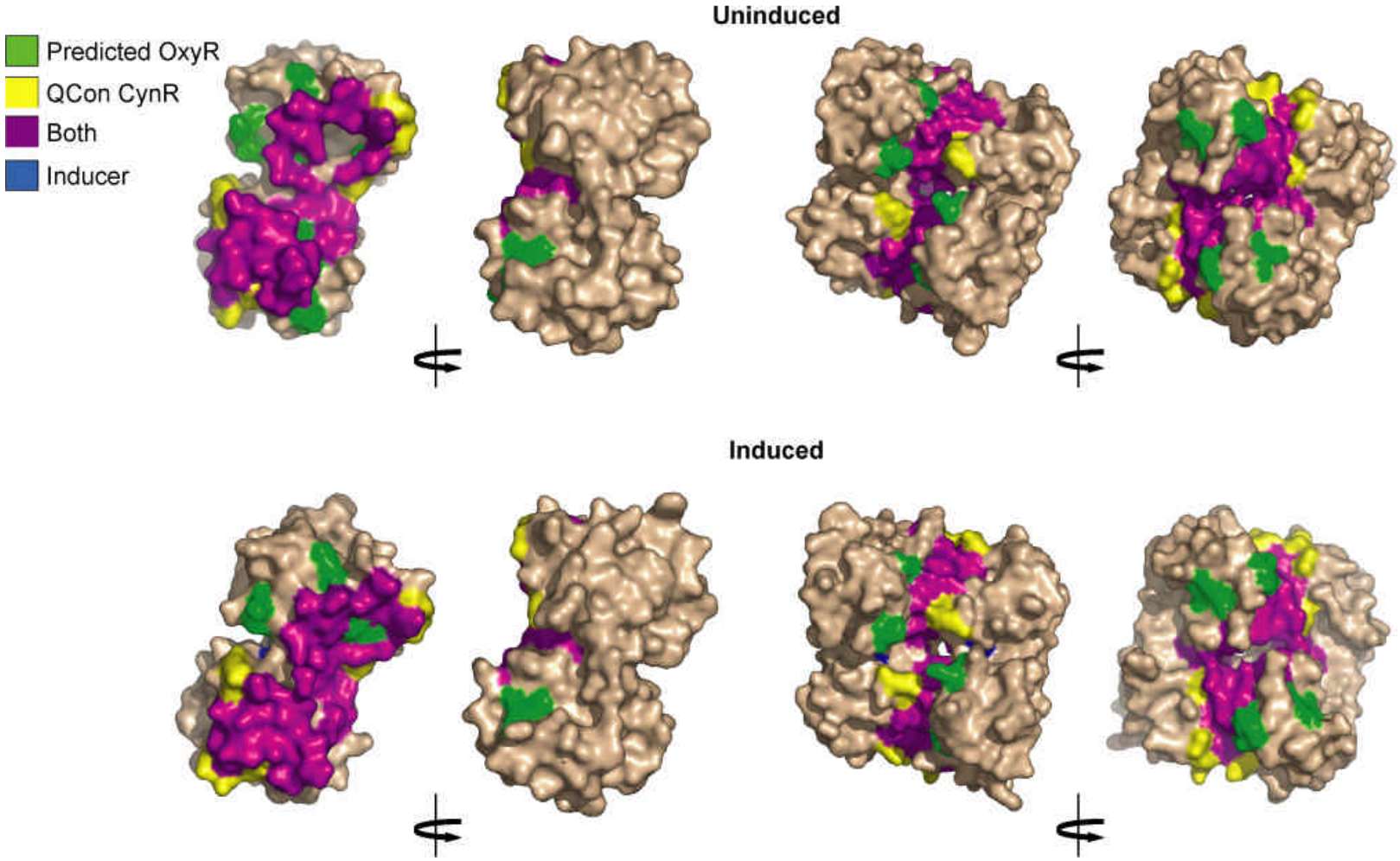


Figure 3.3. Mapped Comparison of the Predicted and Real Contact Profiles of CynR The left-hand side shows one monomer, turned about the y-axis 180°. On the right-hand side are the dimer complexes. Top: Uninduced. Bottom: Induced.

pAZ299 and, in some cases, pGK751. These three vectors contain the λcI with the oligomerization domain replaced by a Gateway cloning cassette. pAZ299 and pLM100 are under the control of a weak, constitutive promoter, P_{7107}. pGK751 is under the control of P_{lacuv5}. Both pGK751 and pLM1000 contain an amber mutation at codon 103 that is suppressed in JH787. These three vectors were chosen to give an increasing amount of fusion protein. Cross-streaks were performed to assay the fusion's ability to oligomerize. In all three vectors, full-length wild-type CynR is immune. We sorted the immunity results of the full-length CynR mutations into three classes: 1) those that were not able to achieve immunity in any of the vectors, 2) those that were unable to oligomerize in pLM1000, but able to oligomerize in pAZ299 or pGK751 and 3) those that behaved as wild-type, being immune in all three vectors. These results are listed in Table 3.8. There are four residues in Class I, sixteen in Class II and nine residues belonging to Class III.

Because residues 89-299 are sufficient to drive oligomerization, and because the tetrameric contacts of the full-length could be masking the destabilization effects of the alanine mutation at the dimeric interface, entry clones of residues 89-299 containing the mutations were constructed. These entry clones were used to construct various λcI fusions with the regulatory domain of CynR. These constructs were assayed for immunity. These results were sorted based on the ability to oligomerize. The results are shown in Table 3.8.

Four residues were found to be required for the oligomerization of full-length CynR as assayed by λcI repressor fusions. We mapped these residues onto the surface

of the structure and examined the residues that were exposed to the surface (Figure 3.4). By simple examination of the atom's exposure to the surface, exposure to the surface, L228 was almost completely buried in both forms of CynR. Because it is not at the surface, we conclude that L228 is not directly involved in oligomerization. However, it could be influencing another residue at the interface. Thus, we conclude that three residues in the dimeric interface of CynR are required for oligomerization: L122, E220 and E229.

Attempts to Detect Oligomerization in the Presence of Cyanate

The small-molecule associated with CynR is cyanate. A general characteristic of LTTRs is that in the absence of inducer molecule, the protein can form its tertiary complex and bind DNA. Upon binding of the inducer, there is a conformational change among the regulatory domains. This conformation change is most readily seen in the OxyR crystal structures, with the reduced and oxidized forms having a 30° rotational change between the subunits upon exposure to oxidative stress (Choi et al. 2001). We wanted to know if immunity of CynR would change in the presence of KCON, the small-molecule associated with CynR. Class I and Class III's ability to oligomerize was assayed on freshly made tryptone plates containing 1 or 10mM KCON. No change in immunity was detected (data not shown).

Table 3.8. Oligomerization Results of CynR

	Full-length	**Regulatory Domain**	**Sub-Class**
L122A	I	I	1A
L228A	I*	I	1A
E229A	I	I	1A
L124A	II	I/II	2A
Q247A	II	I/II	2A
Y104A	II	II	2B
I120A	I/II	II	2B
I121A	II	II	2B
L230A	II	II	2B
R233A	II	II	2B
S235A	II	II	2B
Y114A	II	III	2C
E126A	II	III	2C
Q123A	III	II	3A
Q125A	III	II/III	3A
H172A	III	III	3C
S223A	III	III	3C
E220A	I	ND	
R94A	II	ND	
L106A	II	ND	
M110A	II	ND	
K131A	II	ND	
V218A	II	ND	
R232A	II	ND	
H248A	II	ND	
S103A	III	ND	
S128A	III	ND	
I219A	III	ND	
S225A	III	ND	
T234A	III	ND	

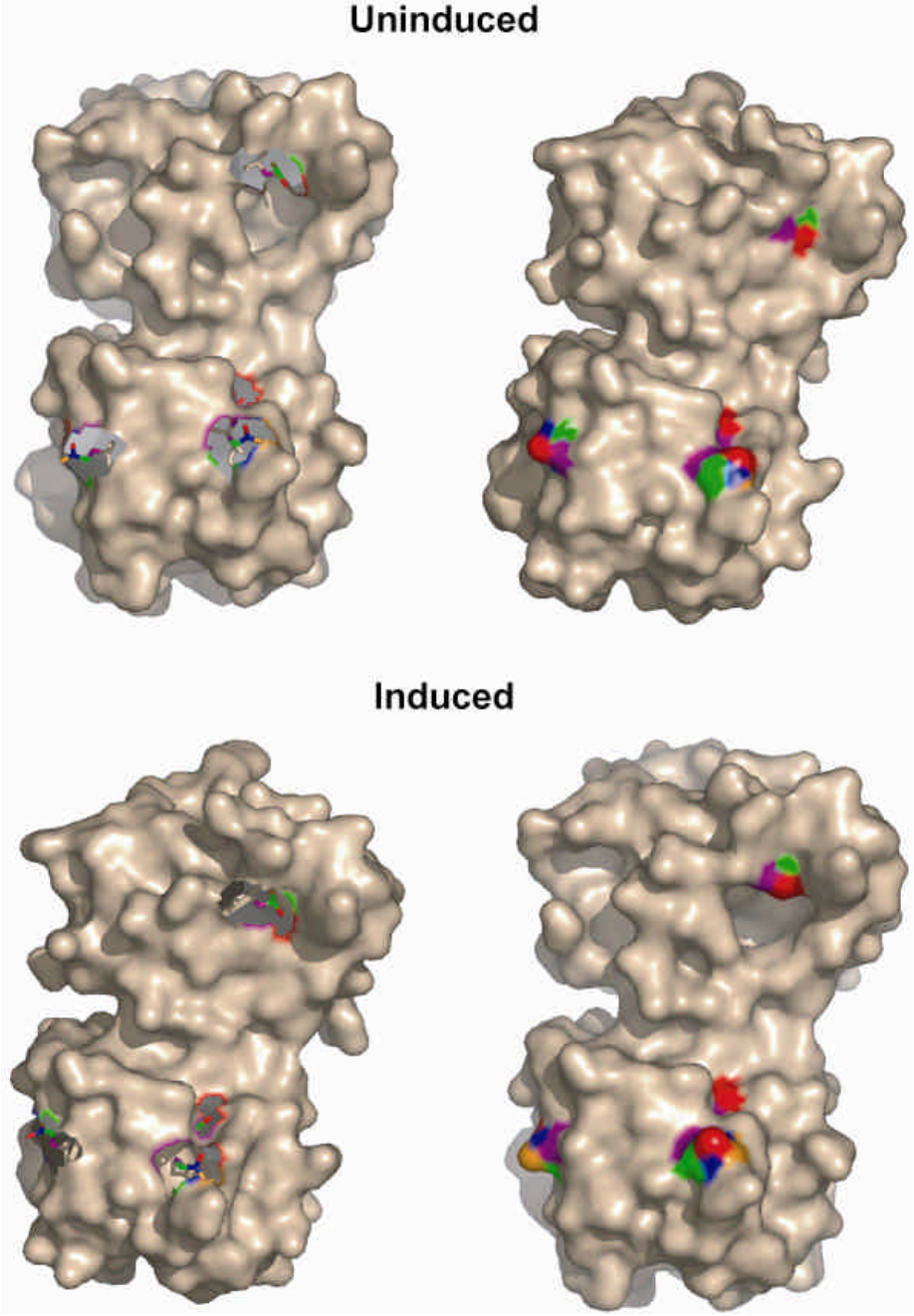


Figure 3.4. Determination of Buried Atoms. On the left-hand side, the surface of one monomer is shown and the hot spots are shown as stick representations. The atoms distal to Ca are colored orange and red. On the right-hand side, the complete surface is shown. Top: Uninduced. Bottom: Induced.

The Search for a Phenotype of a *cynR* Strain

One of the goals of this project was to determine whether or not the λcI-full-length CynR fusions were in fact behaving as CynR. A mutation that did not affect oligomerization could be affecting other functions of the protein. To address this, I attempted to find a growth phenotype for a strain deleted for *cynR*. Strains deleted for either *cynR, cynS* or *cynT* were obtained from the ASKA collection, as well as the parental strain BW2511 (Baba et al. 2006). Growth curves were done to determine if there was a different growth phenotype in response to the addition of KCON. Figure 3.5 shows a representative curve for all four strains at 0, 1 and 5mM KCON. At 1mM KCON, all strains grew at approximately the same rate as the cultures with no KCON. At 5mM KCON, though growth was significantly slower then 0 and 1mM KCON, the OD_{600} did begin to increase around the 150 minute point.

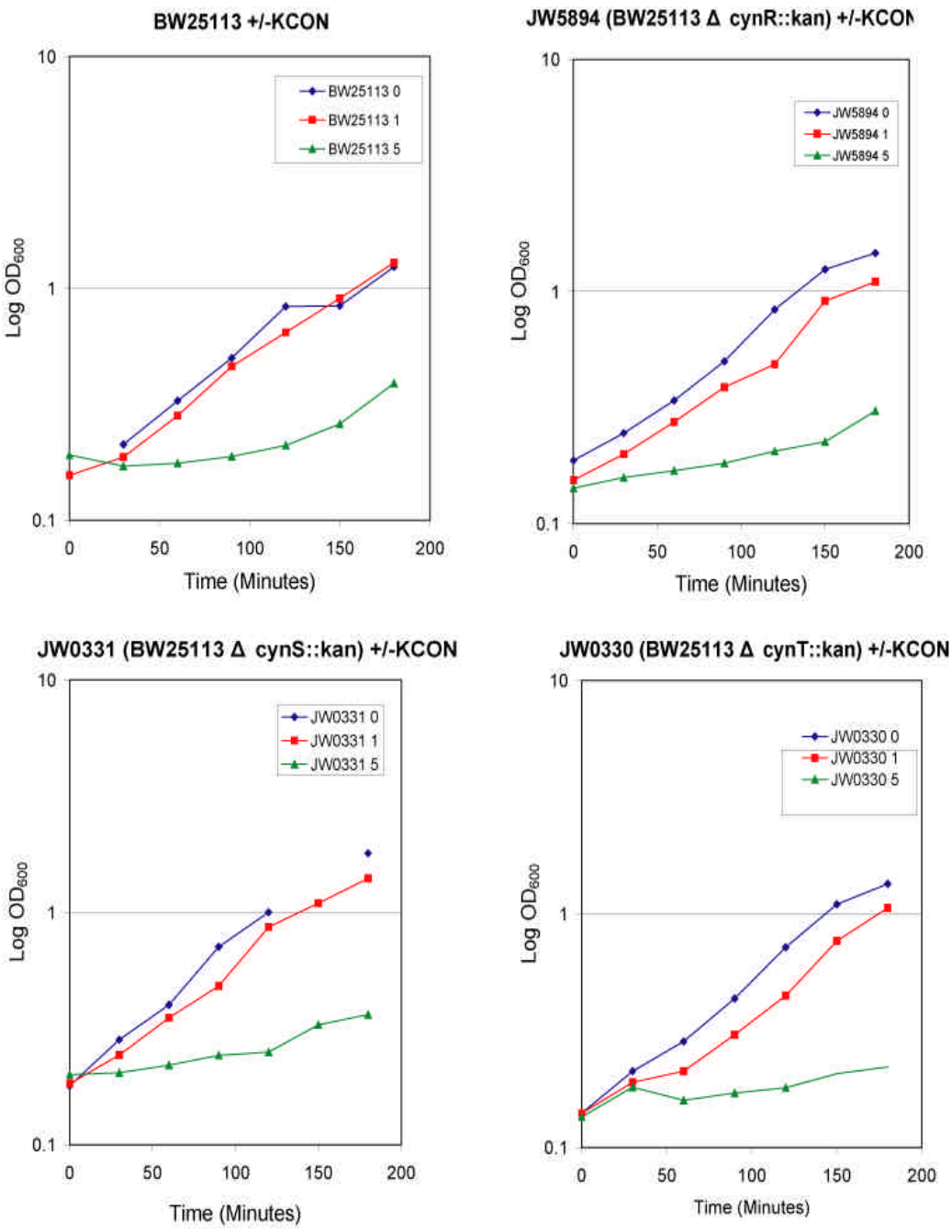


Figure 3.5. Growth Curves of *cyn* Strains

DISCUSSION

Numerous studies of LTTRs, including CysB, NahR and GcvA identify the C-terminal region of the LTTRs participating in oligomerization of the protein (Schell et al. 1990; Jourdan and Stauffer 1998; Lochowska et al. 2001). However, the crystal structure of the full-length protein CbnR suggests that several other regions of the protein could contribute to the oligomerization of the LTTR, including the coiled-coil linker between the DNA binding domain and the regulatory domain, and a subdomain near the C-terminus (Muraoka et al. 2003). Further, mutations in the Leu41-Ile48 region of CysB DNA binding domain of CysB lacked a dominant-negative effect on the chromosomally produced CysB, suggesting that, among other possibilities, this region contributes to oligomerization REF.

By expressing different pieces of CynR as λcI fusion proteins in *E. coli,* we could detect oligomerization domains. We were able to identify domains sufficient for oligomerization, as well as domains that contributed to the stability of the multimer. Truncations containing portions of the regulatory domain were immune to phage infection, indicating oligomerization. In an assay based on cooperative binding to multiple λ operators, these behaved as dimers, while the full-length protein behaved as a higher ordered oligomer, consistent with the observation that most LTTRs are tetramers and clarifies the oligomerization state of CynR. Previous reports by the Fuchs group have said that CynR is a dimer of dimers or it is dimeric as indicated by gel-filtration studies (Anderson et al. 1990; Lamblin and Fuchs 1994).

Constructs D, F and G are sufficient for oligomerization, indicating that residues 88-257 are sufficient for oligomerization. None of these constructs are indicative of higher oligomer states. The region including residues 258-299, which is the only difference between construct D and E, is most likely involved in the stability of the multimer.

The region containing the coiled-coil observed in CbnR in CynR was neither sufficient for driving oligomerization, nor were regions containing the coiled-coil able to form higher-ordered oligomers. Muraoka et al. observed a hydrophobic patch between residues I71, L74 and A75, as well as two hydrogen bonds between residues R81 and D67 and S82 and A60 (Muraoka et al. 2003). Analysis with QContacts detected additional interactions within this region. (data not shown). Assembly of the coiled-coil in the tetrameric protein is most likely driven by the local environment.

We identified three residues important that contribute to the oligomerization of CynR: L122, E220 and E229. Comparisons of these hot spots with the general trends discussed in Chapter I show that the hot spots in CynR share several of the typical characteristics of hot spots. For example, as shown in Figure 3.6, the hot spots make contact with another hot spot. Additionally, the hot spots are surrounded by residues that are not especially important for oligomerization. However, one major difference observed between the hot spots and the characterized hot spots is composition. Studies by Bogan and Thorn and Hu et al emphasize that the hot spots are enriched for tryptophan, tyrosine and arginine (Bogan and Thorn 1998; Hu et al. 2000). Two hot spots are glutamic acid and the other is leucine.

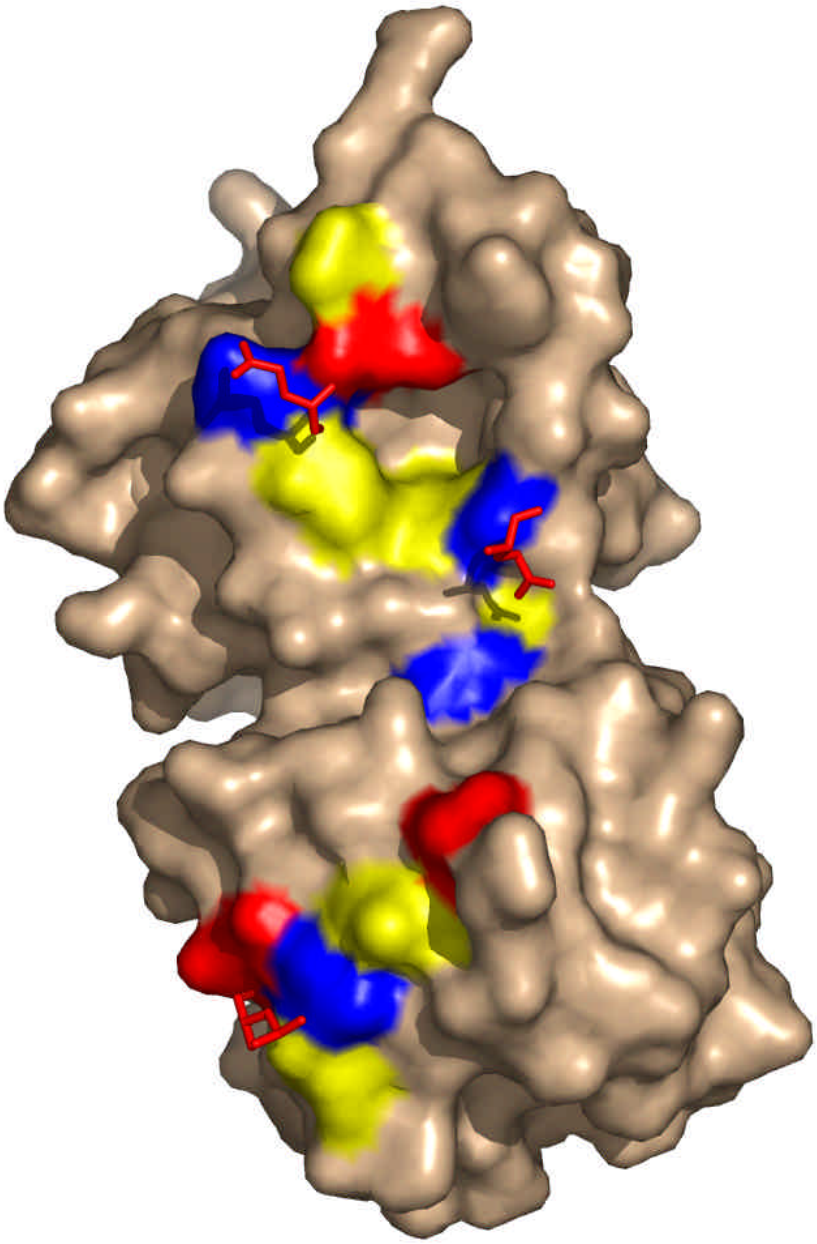


Figure 3.6. Hot Spot Contacts. Above is the uninduced CynR interface. The surface of one monomer is shown. Hot spots are red. Blue residues are Class III residues that are in contact with the hot spot. Yellow are Class II residues. The opposite monomer is shown with the stick representation.

We identified the hot spots through the use of alanine-scanning mutagenesis and these residues were identified as targets through the use of predicted interactions. We used the contact profiles identified in OxyR (Chapter II) as a template to generate a predicted profile of the residues involved in oligomerization in CynR. There was an overlap in the residues in the predicted contact profiles and the contact profiles generated from the crystal structures of CynR. The differences can be attributed to the uniqueness of the OxyR and CynR structures, as shown in Figure 3.3.

Due to the structural similarities between the regulatory domains of the LTTRs, we originally hypothesized that there would be some overlap between the residues identified as hot spots in OxyR and CynR. However, as shown in Figure 3.7, this is not the case. Of the seven residues from the OxyR study, three were not found in either of the contact profiles generated from CynR (Table 3.7). During this study, additional regulatory domain crystal structures were released of various LTTRs (Chapter I). We were able to utilize QContacts on these structures to determine the contact profiles. The contact profiles are mapped in blue in Figure 3.7. While there are some residues that are in the contact profile of all the RD structures, very few overlap with the hot spots of CynR and OxyR, supporting the idea that the protein-protein interactions in the cell are highly specific.

If the hot spots are not shared amongst proteins in the same family, then one could hypothesize that contact residues that are unique to a structure might hold clues to the residues responsible for oligomerization. For example, positions 190, 195 and 197 were identified as being contact residues in the CynR contact profiles, while 170 and 172

were unique contact residues in OxyR. Neither set of residues were found in the other protein. Though 190, 195 and 197 were not tested as mutations, 170 and 172 were and neither contributed to the oligomerization of OxyR nor to the oligomerization of CynR.

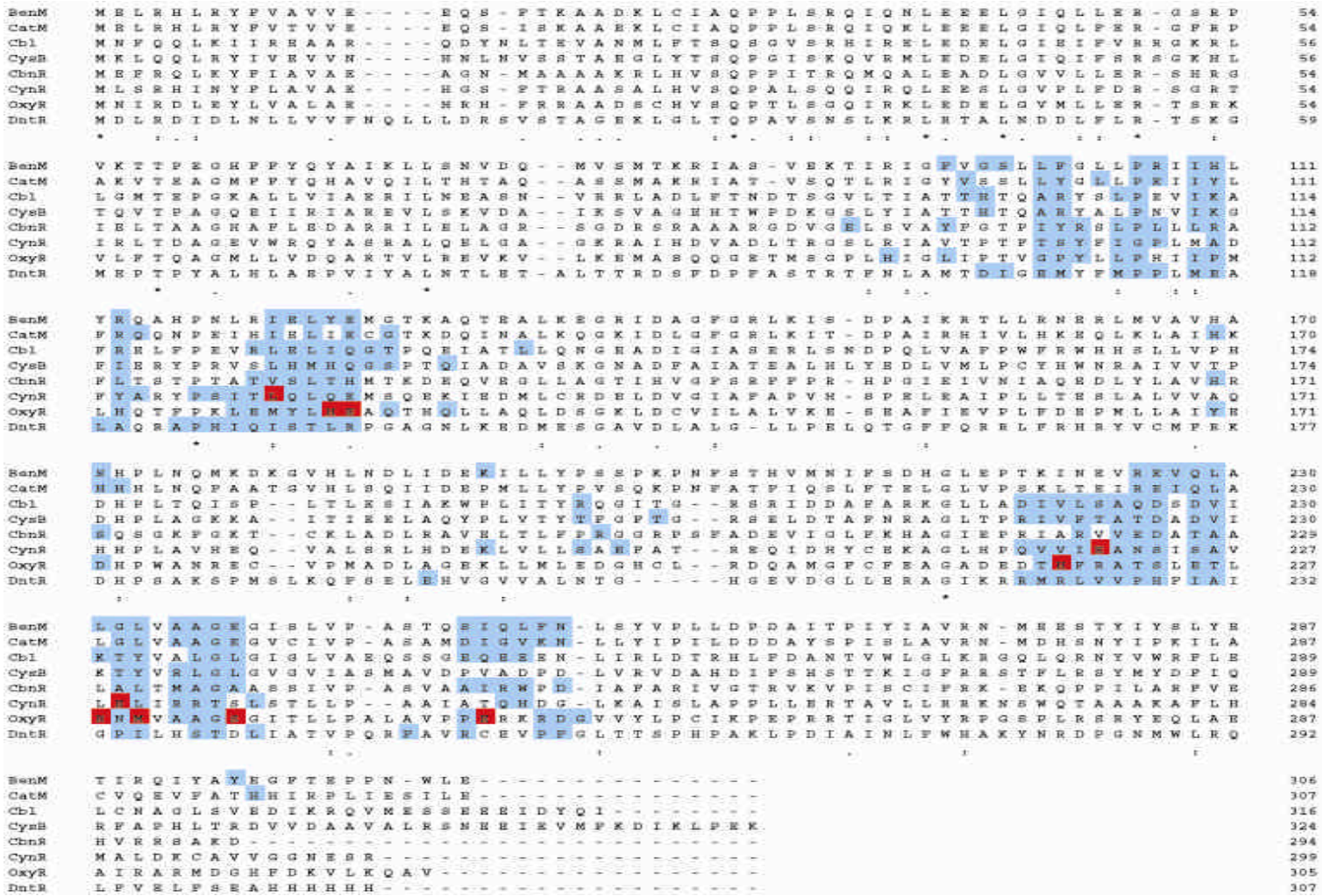


Figure 3.7. Contact Residues and Hot Spots Mapped onto the Multiple Sequence Alignment of the LTTRs with a Crystal Structure.

CHAPTER IV

HETEROTYPIC INTERACTIONS OF THE LysR-TYPE TRANSCRIPTIONAL REGULATORS

OVERVIEW

How the protein knows which protein to form a partner with is a fundamental question of biology. The LysR-Type Transcriptional Regulators form perhaps the largest family of transcriptional regulators. Because genomes often encode many LTTR family members, it is assumed that many distinct oligomers are formed simultaneously in the same cell without interfering with each other's activities, suggesting specificity in the interactions. However, this assumption has not been tested. Through the use of a negative-dominant assay with λcI repressor fusions, I show that the LTTRs in *E. coli* K-12 are mostly specific in their interactions.

INTRODUCTION

The LysR-Type Transcriptional Regulators (LTTRs) are a diverse family of oligomeric transcription factors that are found in prokaryotes. The family was initially identified by Henikoff in 1988 with roughly 50 members (Henikoff et al. 1988). With the continuous sequencing of new genomes, that number has grown to over 18,000 potential members (IPR000847 HTH_LysR (Quevillon et al. 2005)), making it perhaps the largest family of transcriptional regulators among prokaryotes. Numerous LTTRs are found in a single species. For example, *E coli* K-12 has 47 members, while

Pseudomonas aeruginosa has 112 members (Tatusov et al. 2001).

Studies investigating the oligomerization of the LTTRs have demonstrated that they are homooligomeric proteins, consisting mostly of tetramers and in some cases, dimers. As reviewed in Chapter I, examination of the available crystal structures shows an overall structural similarity (see Figures 1.2 and 1.3). These interactions are essential for function of the protein. With the abundance of proteins in the cell, it could be potentially detrimental to the cell if the LTTRs could form heteromultimers, suggesting that complex formation should be highly specific. However, the specificity of the LTTRs has not been explicitly tested. In this study, we utilize the available λcI repressor LTTR fusions in a negative-dominance assay to test for the LTTR's ability to form heterotypic interactions.

METHODS

Construction of Fusions

Immune λcI LTTR fusions were identified in a screen for homotypic interactions in *E. coli* K-12 (Marino-Ramirez et al. 2003). Thirteen λcI LTTR fusions were identified. To take advantage of the Gateway recombination system, the clones were amplified from the plasmid DNA using primers *att*B1 and *att*B2 (PAGE purified, IDT, Iowa) which attach Gateway cloning sites. Reactions were done in 50µl total volume, using Pfx Platinum (Invitrogen). The size of the amplified product was checked on a 1% TBE Agarose gel and a Qiagen PCR Clean-up Kit used.

Entry clones were generated using 2µl of the eluted PCR product, 2µL of the 5x BP Clonase Buffer (Invitrogen), 2µl TE, 2µl of 100 ng/µL pDONOR201 and 2µL BP Clonase (Invitrogen). The reactions were incubated together overnight at room temperature and transformed into either MC1061 or Mach T1 (Invitrogen). In addition to the constructs described above, we obtained the rest of the LTTRs cloned as λcI repressor fusions. Entry clones were generated *via* the back reaction by Adrienne Zweifel. Relevant clones are described in Table 4.1.

Negative-Dominant Assay

Zeng *et. al* (Zeng et al. 1997) developed a negative-dominance assay based on the λcI repressor system in which two constructs are used and is illustrated in Figure 4.1. The first, termed termed cI^+, is a λcI fusion with a wild-type DNA binding domain with a protein fused to it that is stable enough to oligomerize and confer immunity to λ infection. This fusion is expressed from a weak, constitutively expressed promoter, The second construct has a mutated cI DNA binding domain that is unable to bind DNA and is expressed in an excess amount in the cell.

To test for heterotypic interactions using this dominant-negative assay, strains carrying the co-expressed plasmids are exposed to phage λ. Fusions that can form heterotypic interactions will be sensitive to λ if the cI^+ homooligomer levels fall below the critical level required for immunity to λ. Immunity to λ is scored when the cI^+/ cI^+ complex is more stable than a cI^+/cI^- complex. For the negative-dominant fusion, we utilized a thioredoxin- fusion protein. The thioredoxin plasmid, pJM198, is a modified

Table 4.1. Relevant Strains and Plasmids

					pLM1000		pJM197	
Gene Name	**IST ID***	**Location**	**GK Name**	**IST in pDONR201**	**MC1061**	**JH787**	**MC1061**	**JH787**
yhjC	EB099.012	EB09901H12	GK300	GK313	GK326	GK339	GK352	GK365
yafC	EH099.013	EH09901F01	GK303	GK316	GK329	GK342	GK355	GK368
cynR	EH100.049	EH10002E11	GK304	GK317	GK330	GK343	GK356	GK369
pssR	EH100.093	EH10001H02	GK305	GK318	GK331	GK344	GK357	GK370
yiaU	EH100.118	EH10004A06	GK306	GK319	GK332	GK345	GK358	GK371
gcvA	EH100.160	EH10004H12	GK307	GK320	GK333	GK346	GK359	GK372
ilvY	EH101.053	EH10102F02	GK308	GK321	GK334	GK347	GK360	GK373
ynfL	EH101.073	EH10103H12	GK310	GK323	GK336	GK349	GK362	GK375
iciA	EH101.093	EH10103C12	GK312	GK325	GK338	GK351	GK364	GK377
NhaR							RS322	RS345
YneJ	pMD33						RS321	RS344
YgiP	pMD28			AZ488			RS320	RS343
YdhB	pMD22			AZ482			RS318	RS341
YbeF	pMD16			AZ478			RS316	RS339
YbbS	pMD14			AZ477			RS315	RS338
YagP	pMD12			AZ475			RS314	RS337
TdcA	pMD10			AZ473			RS313	RS336
LysR	pMD5			AZ468			RS312	
YfiE	pMD26			AZ486			RS311	RS334
YidZ	pMD29			AZ490			RS310	GK385

Table 4.1. Continued

					pLM1000		pJM197	
Gene Name	**IST ID***	**Location**	**GK Name**	**IST in pDONR201**	**MC1061**	**JH787**	**MC1061**	**JH787**
YgfI	pMD27			AZ487			RS309	RS332
YdaK	pMD20			AZ481			RS308	RS331
YbhD	pMD17						RS307	RS330
YahB	pMD13			AZ476			RS306	RS329
PerR	pMD9			AZ472			RS305	RS328
OxyR	pMD8			AZ471			RS304	RS327
MetR	pMD6			AZ469			RS303	RS326
LrhA	pMD4						RS302	RS325
Cbl	pMD1			AZ465			RS300	RS323
CysB	pMD2			AZ466			RS301	RS324
GCN4				GK210		LM239	JM185	GK221

*Strains starting with an E are from the IST Screen and are located in microtiter plates (Marino-Ramirez et al. 2004). I made freezer stocks and gave the clones GK names. The plasmids listed as pMD*** were obtained from Hirotada Mori as λcI repressor fusions cloned into *Sfi*Isites. They were obtained by Adrienne Zweifel and are in her collection. Note, however, that entry clones with AZ numbering *do not* have a stop codon and most, if not all, were sensitive as λcI fusions without the stop codon.

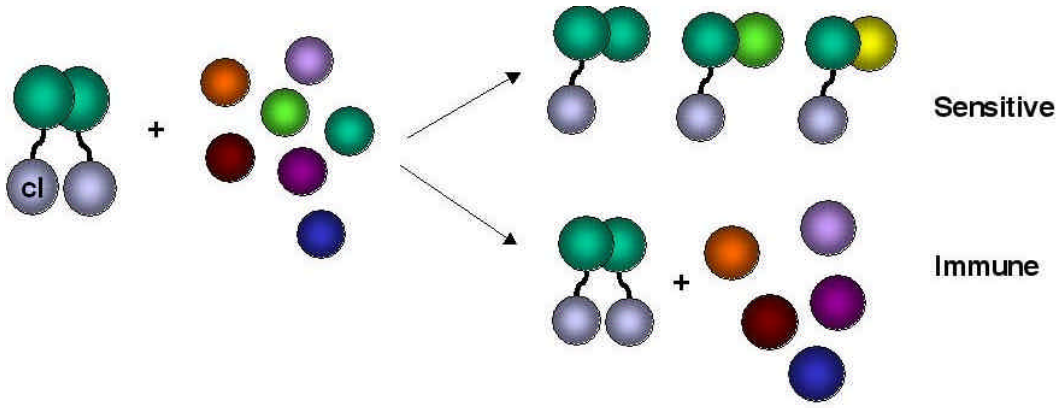


Figure 4.1. Negative-dominance Assay. The oligomerizing cI fusion is shown on the left-hand side. The proteins tested for heterotypic interactions are represented by the colored spheres. (Top, right) Proteins that can form heterotypic interactions with the immune cI fusion generate a sensitive phenotype. Those that can not form an interaction remain immune.(Bottom, right)

version of pBAD49 and made by Jon Minor and Gregory Munoz. The backbone was swapped out to have pACYC184 origin, as well as the tetracycline resistance gene. Entry clones described above were moved into this construct *via* the LR reaction.

Construction of Double Strains

A general overview of the double strain construction is shown in Figure 4.2. The double strains were constructed using M13 transduction. M13 RV-1 was used to pack the immune λcI repressor fusions. Cells were grown in 2XYT and ampicillin (200μg/mL). The strains containing the thioredoxin fusions were grown in 2XYT containing tetracycline (20μg/mL). Using a Pasteur pipette, two drops of the thioredoxin fusion-containing strains were put into a sterile 96-well plate. 5μL of the packed M13 strain was put into the appropriate well (typically using a multichannel pipette). 100μL of LB was added to each well after incubation for 10 minutes at room temperature. Airpore tape was put on top of the plate and incubated at 37°C for 120 minutes.

For selection of the double-strains, a 96 well frogger was used to stamp LB-ampicillin-tetracycline plates. The plates were incubated overnight at 37°C. The next day, the entire grid was stamped onto fresh LB-ampicillin-tetracycline plates and incubated again. This was considered streak purification or stre-frogging. Cells were selected from this spot for cross-streaks. No freezer stocks were constructed.

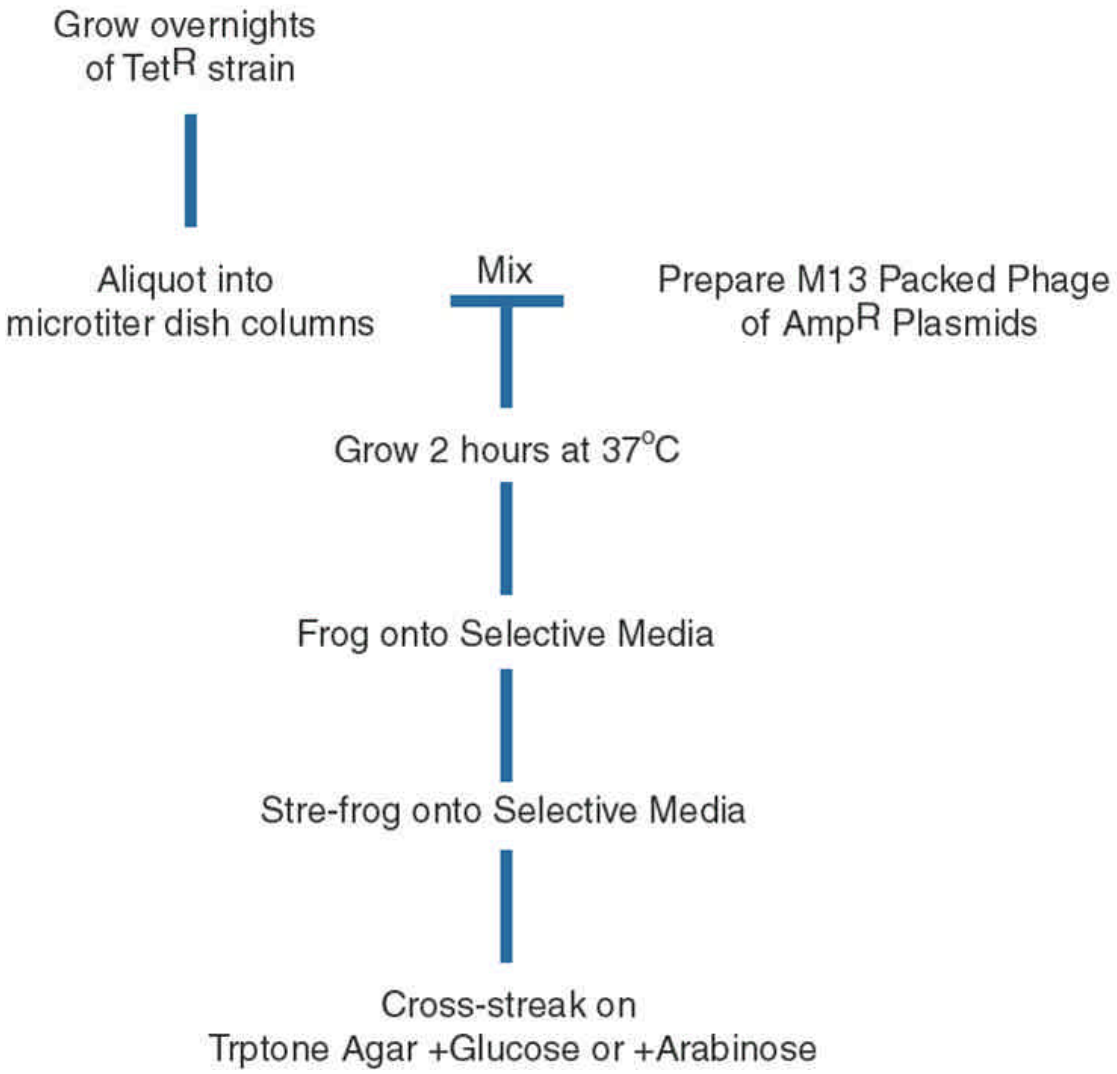


Figure 4.2. Overview of Double Strain Construction.

Cross-streak Assays

Cross-streak assays were done to assay the constructs' ability to oligomerize. Strains from the streak purification were tested on plates containing tryptone or tryptone containing 0.2% glucose or tryptone 0.2% arabinose. Cells were first touched to the plate that did not contain arabinose, then to the arabinose plate. A fresh toothpick was used to draw the cells across the lines of phage (λKH54 and λ^{i2}c). Plates were incubated at 37°C and the phenotype noted. For these large scale interactions, the same plates of colonies were used to test on three subsequent days.

RESULTS

Reconstitution of the Negative-Dominance System with Thioredoxin Fusions

We initially constructed a small grid of λcI LTTR repressor fusions with their thioredoxin counterparts. Thioredoxin GCN4 and λcI GCN4 was used as a control. This small set of interactions is outlined in the blue box in Figure 4.3. The thioredoxin GCN4 did not interact with any repressor fusions except λcI GCN4. λcI GCN4 did not interact with any other thioredoxin fusion except GCN4.

All λcI LTTR proteins interacted with their respective thioredoxin fusion, as illustrated by the plum colored boxes in Figure 4.3. A few heterotypic interactions were observed, particularly with the trx-PssR fusion agsinst λcI-YafC, λcI-CynR and λcI-YiaU.

Expansion of the Grid

We next used the λcI LTTR fusions to test against 23 other trx-LTTR fusions. The resultant interactions are shown in Figure 4.3. Several trx-LTTR fusions had numerous interactions with the λcI LTTR fusions, including trx- Cbl, OxyR, PerR and YbeF. Others, such as LysR, NhaR and YneJ showed only one or two interactions. Many showed no interactions with the λcI LTTR fusions.

Results of Both Grids

Of the 236 potential heterotypic interactions, 38 interactions were detected, allowing for three classes of interactions: I) those LTTRs that did not interact with other, II) LTTRs that formed only a few interactions and III) LTTRs that were promiscuous in their interactions with other LTTRs. LTTRs in the promiscuous category are Cbl, OxyR, PerR, YbeF and YdaK. Class II includes LTTRs CynR, LysR, NhaR, YneJ and YiaU; all other LTTRs fall into Class I.

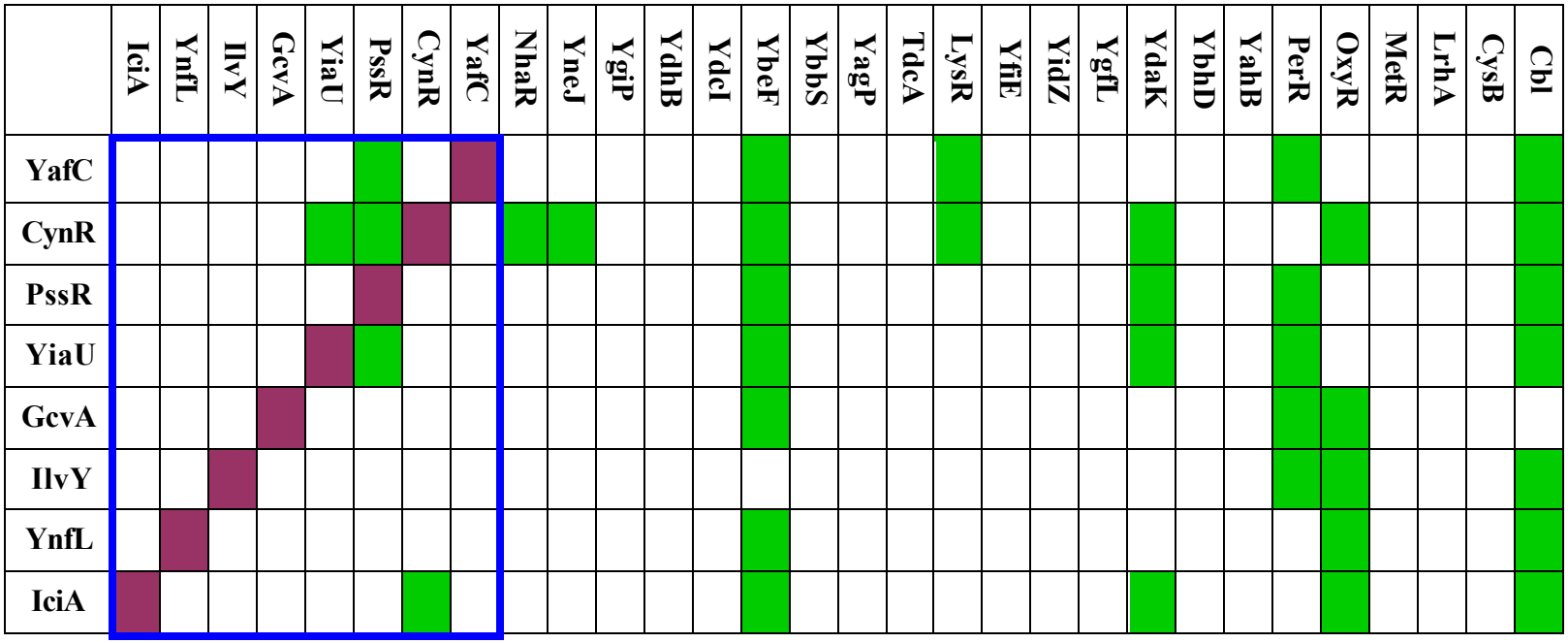


Figure 4.3. Interaction Grid of the LTTRs. On the left are the immune λcI repressor fusions. Across the top are the trx-LTTR fusion constructs. Displayed is the result of 1 colony tested on three different days. An interaction was noted if at least two of cross-streaks generated a sensitive phenotype. Here, an interaction between two proteins is denoted by a colored square. Plum is representative of a homotypic interaction, while green is a heterotypic interaction. White implies no interaction.

DISCUSSION

In this work, we present evidence that the LTTRs are specific for the formation of homotypic interactions, as demonstrated by the pair-wise testing of 236 potential heterotypic LTTR interactions. These interactions, which we have grouped into three classes, show that Class I is the largest class. Class II and Class III represent LTTRs that are able to form heterotypic interactions.

Closer examination of the interaction patterns of Class II and Class III does not reveal a clear-cut interaction pattern and rules of interaction can not be generated from the interaction patterns in Table 4.2. The Class III proteins do not interact with all the LTTRs, nor do they interact with only certain LTTRs. The same can be said for Class II proteins. While they only interact with one or two other LTTRs, the interactions appear to be random.

If the interactions are true, then these LTTR might represent LTTRs that are not expressed at the same time. Thus, the pressure to have evolved high specificity might be less than that of other LTTRs. However, we have not found any evidence that LTTRs are *not* expressed at the same time. In fact, we would expect all LTTRs, though at low levels of LTTRs, to be expressed at the same time, as most LTTRs repress themselves. Thus, this idea supports the notion that the LTTRs are highly-specific. Therefore, we must consider that the Class II and Class III proteins could be artifacts.

The fact that the LTTRs are specific for homotypic interactions fits well with the take home message in Chapter III. The LTTRs that we have studied do not have a shared set of residues important for oligomerization. If each of those residues contributes

to the specificity of the molecule, then the results presented here only further validate our hot spot data. Understanding how these residues have evolved for the specificity of the protein will require further exploration.

CHAPTER V

RESISTANCE TO METHOTREXATE DUE TO AcrAB-DEPENDENT EXPORT FROM *Eschericia coli**

OVERVIEW

Many laboratory strains of *Escherichia coli* are resistant to methotrexate (MTX), a folate analogue that binds dihydrofolate reductase (DHFR). Mutations that inactivate either *tolC* or *acrA* confer MTX sensitivity. Further, overexpression of a fusion protein with DHFR activity reverses this sensitivity by titrating out intracellular MTX. These results suggest that MTX accumulates in cells where mutations in *acrA* or *tolC* have inactivated the TolC-dependent AcrAB multidrug resistance efflux pump.

INTRODUCTION

Methotrexate (MTX) is a folate analogue that inhibits the activity of dihydrofolate reductase (DHFR) (Miller 1972), which catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate. Reduced folates are substrates in a number of one-carbon transfers in purine, pyrimidine, and amino acid biosynthesis (Baccanari et al. 1981).

---

*The work described in this chapter was previously published as : Kopytek, S.J., Dyer, J.C., Knapp, G.S., and Hu, J.C. 2000. Resistance to methotrexate due to AcrAB-dependent export from *Escherichia coli*. *Antimicrob Agents Chemother* **44:** 3210-3212.

Inhibition of DHFR activity initially results in the depletion of N^5,N^{10}-methylene tetrahydrofolate, followed by inhibition of DNA synthesis and ultimately cell death (Howell et al. 1988). DHFR is thus a well-studied target of antibiotic and antineoplastic therapy.

Although MTX binds both human and *Escherichia coli* DHFR very tightly, with K_i values of 3.4 and 1.0 pM, respectively (Appleman et al. 1988), all of the *E. coli* isolates we tested (genotypes of the strains used in this study are listed in Table 5.1), which included both common laboratory strains (MG1655, MC4100, AG1688, and ZK126) and clinical isolates (O157:H7, RM74A, STM1, LL, RM52B, DD, and RM33B), were resistant to MTX added to solid medium at concentrations of up to 1 mM, the highest concentration we tested (data not shown).

Antibiotic resistance can occur by a variety of mechanisms, including failure of the drug to bind its target, overexpression of the drug target, modification or degradation of the drug, creation of permeability barriers, or active export of the drug. It is increasingly recognized that active efflux plays a major role in the resistance of many organisms to a plethora of agents (Levy 1992; Nikaido 1994). A wide variety of antibiotics are exported from *E. coli* by one of several active efflux systems (Levy 1992; Lewis 1994; Nikaido 1994; 1996). At least two of these systems, the AcrAB and EmrAB efflux pumps, have been shown to depend on the outer membrane protein TolC (Lewis 1994; Nikaido 1994; Fralick 1996; Nikaido 1996; Aono et al. 1998).

To determine whether the MTX resistance was due to a TolC-dependent efflux pump, we examined the effect of a *tolC*::Tn*10* mutation. LBB1175, in which *tolC* had been inactivated by the Tn*10* insertion, was sensitive to 1 mM MTX, while W4573, the isogenic TolC$^+$ control, was resistant. Similar results were obtained using the common laboratory strain MG1655, which is the reference wild-type *E. coli* K-12 strain used for the genome sequence (Blattner et al. 1997), and AG1688 (see below). Strains carrying Tn*10* at a different chromosomal location remained resistant to MTX. These results suggest that MTX resistance is mediated by a TolC-containing multidrug resistance efflux pump (MDR).

tolC mutants are pleiotropic (Morona et al. 1983; Wandersman and Delepelaire 1990) and are hypersensitive to many hydrophobic agents (Nagel de Zwaig and Luria 1967). Thus, the loss of MTX resistance in the *tolC* mutant might not be due to the loss of function of an MDR. To address this possibility, we tested the effects of mutations that inactivate specific TolC-dependent MDRs. The AcrAB pump belongs to the RND (for resistance, nodulation, and division) family,

Table 5.1. *E. coli* Strains Used

Strain	Genotype	Reference and/or source
MG1655	K-12 $F^-\lambda^-$	(Perkins et al. 1993); D. Siegele
MC4100	F^- *araD139* Δ(*argF-lac*)*U169 rpsL150 relA1 deoC1 rbsR flhD5301 fruA25* λ^-	(Bremer et al. 1984); D. Siegele
AG1688	F'128 *lacI*^{q} *lacZ*::Tn*5*/*araD*139 Δ(*ara-leu*)7697 Δ(*lac*)*X74 galE15 galK16 rpsL*(Str^r) *hsdR2 mcrA mcrB1*	MC1061 F'128 *lacI*^{q} *lacZ*::Tn*5* (Hu et al. 1993)
ZK126	W3110 Δ(*lac*)*U169 tna-2*	(Dougherty and Pucci 1994); D. Siegele
O157:H7	*E. coli* isolate from human	(Izumiya et al. 1997); D. Siegele
RM74A	Group I *E. coli* from human female	(Milkman 1973); D. Siegele
STM1	Group I *E. coli* from human male	(Whittam et al. 1983); D. Siegele
LL	Group II *E. coli* from human infant	(Selander and Levin 1980); D. Siegele
RM52B	Group II *E. coli* from human female	(Milkman 1973); D. Siegele
DD	Group III *E. coli* from human infant	(Whittam et al. 1983); D. Siegele
RM33B	Group III *E. coli* from human female	(Milkman 1973); D. Siegele
W4573	K-12 F^- *lac ara mal xyl mtl gal rpsL*	(Ma et al. 1993); J. A. Fralick
LBB1175	W4573 *tolC*::Tn*10*	J. A. Fralick
N43	W4573 *acrA1*	(Ma et al. 1993); J. A. Fralick
OLS103	AMS6 *emrB*::Km	(Lomovskaya et al. 1995); J. A. Fralick
AMS6	K-12 F^- Δ*lacU169*	(Lomovskaya et al. 1995); J. A. Fralick
SK627	W4573 *acrA1 emrB*::Km	N43 × P1 *vir* (OLS103)
SK636	W4573 *emrB*::Km	W4573 × P1 *vir* (OLS103)
KH803	MC4100 *tolC*::Tn*10*	R. Young
SK642	W4573 *acrA1 tolC*::Tn*10*	N43 × P1 *vir* (KH803)
SK660	W4573 *acrA1 emrB*::Km *tolC*::Tn*10*	SK627 × P1 *vir* (KH803)

Table 5.1. Continued

Strain	Genotype	Reference and/or source
SK037	AG1688 *tolC*::Tn*10*	AG1688 × P1 *vir*KH803)
SK029	AG1688(pSK029)	pSK029, a pBR322-derived plasmid that expresses a λ cI-*E. coli* DHFR fusion protein from P_{lac}UV5, was introduced by M13-mediated transduction (Vershon et al. 1986) into AG1688
XZ020	AG1688(pXZ020)	pXZ020, a pBR322-derived plasmid that expresses a λcI-GCN4 leucine zipper fusion protein from P_{lac}UV5

and its substrates include sodium dodecyl sulfate, basic dyes, novobiocin, and tetracycline (Nikaido 1994; 1996); the EmrAB pump belongs to the MF (major facilitator) family, and its substrates include carbonyl cyanide *m*-chlorophenylhydrazone, nalidixic acid, and phenyl mercury acetate (Nikaido 1994; 1996). A strain containing the *acrA1* mutation (N43) was sensitive to 1 mM MTX, while its isogenic parent (W4573) was resistant. In contrast, both the *emrB* mutant (OLS103) and its isogenic parent (AMS6) were MTX resistant. These results show that the MTX sensitivity of the *tolC* strains is at least partly due to inactivation of the AcrAB MDR, while the EmrAB pump does not have a major role in MTX export.

MICs of MTX were determined for a set of isogenic *E. coli* strains containing combinations of *acrA*, *emrB*, and *tolC* mutations (Table 5.2). The wild-type strain (W4573) was resistant to 1,024 μM MTX, the highest concentration tested. The *emrB*

mutation did not affect the MIC, either alone (compare W4573 to SK636) or in combination with *acrA1* (compare N43 to SK627) or *tolC*::Tn*10* (compare SK642 to SK660). Inactivation of either *acrA* or *tolC* resulted in a decrease of the MTX MIC to 256 or 64 μM, respectively. Since the *acrA1* allele is an IS*2* insertion in the second codon of *acrA* (Ma et al. 1993), it is unlikely that the remaining MTX resistance in the *acrA1* mutant is due to residual activity of the *acrA* gene product. The *tolC* gene product seems to have more than one role in MTX resistance. It is unclear if this is due to the loss of function of another, unidentified TolC-containing MDR or the highly pleiotropic effects of *tolC* mutations on outer membrane structure (Morona et al. 1983; Wandersman and Delepelaire 1990). Similar alterations have not been found in the outer membrane of *acr* mutants (Nikaido and Vaara 1985; Vaara 1993; Nikaido 1996).

The additional role of TolC is not related to the EmrAB MDR, since the *acrA1 emrB* double mutation (in SK627) yielded an MIC identical to that yielded by the *acrA1* single mutation. Similar combinations of mutations in a different background (AG1688) yielded identical MICs (data not shown). This further demonstrates that the observed effects are not strain specific.

Table 5.2. Sensitivities of *acrA*, *tolC*, and *emrB* Mutants to MTX

Strain	Relevant genotype[a]	MIC[b] of MTX (μM)
W4573	Wild type	>1,024
SK636	*emrB*	>1,024
N43	*acrA1*	256
SK627	*acrA1 emrB*	256
LBB1175	*tolC*	64
SK642	*acrA1 tolC*	64
SK660	*acrA1 tolC emrB*	64
[a] For a full listing of genotypes, see Table 5.1.		
[b] MICs were determined by examining the growth of 2-ml liquid cultures containing twofold serial dilutions of MTX (2 to 1,024 μM) in Luria-Bertani liquid broth. The inoculum (20 μl) contained approximately 10^5 cells per ml. The MIC was determined as the lowest concentration that prevented visible growth after 8 h on a roller drum at 37°C. All the tested strains grew to saturation in the absence of MTX. The values are the averages of three separate experiments.		

To determine whether inhibition of DHFR was sufficient to explain the MTX sensitivity of *tolC* strains, we examined whether the MTX sensitivity of AG1688 *tolC*::Tn*10* (SK037) could be suppressed by overexpression of DHFR activity. In the course of other (unpublished) studies, we had constructed a plasmid, pSK029, which expresses a fusion protein, cI-DHFR, in which the N-terminal DNA binding domain of the bacteriophage λ repressor is fused to *E. coli* DHFR; the fusion protein is expressed under the control of the *lac*UV5 promoter. Neither pSK029 nor pXZ020, a control plasmid expressing cI-GCN4 (a fusion to the leucine zipper of GCN4), affected the MTX resistance of wild-type AG1688 whether or not the fusion proteins were overexpressed (Table 5.3, lines 3 and 5). AG1688 *tolC*::Tn*10* strains containing either plasmid were

sensitive to MTX under conditions in which the fusion proteins were uninduced (Table 5.3 lines 4 and 6). However, in the presence of isopropyl-ß-D-thiogalactopyranoside (IPTG), which induces the overexpression of cI-DHFR, SK029 *tolC*::Tn*10* was resistant to high concentrations of MTX (Table 4.3, line 4). In contrast, IPTG-induced overexpression of the control protein cI-GCN4 had no protective effect on the *tolC* strain (Table 5.3, line 6).

These results show that MTX sensitivity in the *tolC* strain is due to the inhibition of endogenous DHFR by the drug. When cI-DHFR is overexpressed, the DHFR activity provided by the DHFR domain in the fusion protein cannot be titrated out, which strongly suggests that the plasmid-coded DHFR acts to sequester MTX that is added to the medium. Increasing the level of DHFR should not relieve sensitivity due to mechanisms that do not involve uptake of MTX.

The results of this study can be summarized as follows. (i) All of the TolC$^+$ AcrA$^+$ strains of *E. coli* we tested were resistant to at least 1 mM MTX when grown on solid medium containing the drug. (ii) MTX resistance is decreased by mutations that disrupt *tolC* or *acrA*, genes that code for integral components of the AcrAB MDR, suggesting that resistance is due to active export of MTX via the AcrAB MDR. (iii) Mutation of the *emrB* gene does not decrease MTX resistance, suggesting that MTX is not a substrate of this MDR. (iv) The difference between the MICs for *tolC*::Tn*10* and *acrA*::IS*2* strains suggests the possibility of another mechanism for low-level TolC-dependent MTX resistance.

Table 5.3. Suppression of the MTX Sensitivities of TolC Mutants by cI-DHFR

Line	Strain[b]	Fusion protein	*tolC* genotype	Growth[a]			
				−IPTG		+IPTG	
				-MTX	+MTX	-MTX	+MTX
1	AG1688	None	*tolC*$^+$	+	+	+	+
2	SK037	None	*tolC* mutant	+	_	+	_
3	SK029	cI-DHFR	*tolC*$^+$	+	+	+	+
4	SK029 *tolC*::Tn*10*	cI-DHFR	*tolC* mutant	+	_	+	+
5	XZ020	cI-GCN4	*tolC*$^+$	+	+	+	+
6	XZ020 *tolC*::Tn*10*	cI-GCN4	*tolC* mutant	+	_	+	_

[a] Cultures were grown to saturation in Luria-Bertani broth at 37°C overnight and diluted in M9 salts to approximately 5,000 CFU/ml. A total of 10 µl of each diluted culture was pipetted onto Luria-Bertani agar plates containing no IPTG (−IPTG) or 1 mM IPTG (+IPTG) and either no MTX (−MTX) or 1 mM MTX (+MTX) as indicated. The spots were allowed to dry and then the plates were incubated at 37°C overnight. +, growth; −, no growth. The observations are from at least three separate experiments.

[b] tolC::Tn*10* was introduced into the indicated strains by P1 *vir* transduction using KH803 as the donor.

CHAPTER VI

CONCLUSIONS

The driving question behind this dissertation is "What is necessary for oligomerization in the LysR-Type Transcriptional Regulators?" Understanding the nature of this oligomerization involved determination of which proteins interact and also characterizing a residue's contribution to oligomerization in individual LTTRs.

Specificity lies in the protein's ability to discriminate between related molecules and to identify the appropriate target molecule. In Chapter IV, I demonstrated that the LTTRs from *E. coli* are specific for the formation of homotypic interactions and the consequences of not forming the proper interactions are discussed in that chapter. How, though, is that specificity determined? To try and address that question, we sought to identify the residues that were important for oligomerization in two LTTRs.

Our approach was to determine the residues in contact at the oligomeric interface and use alanine-scanning mutagenesis to test the contribution of each interface residue in driving oligomerization through the use of λcI repressor fusions. As described in Chapter II, the seven hot spots from OxyR are not especially conserved. Among the OxyR orthologs, S235 was the least conserved of all the hot spots, with a conservation of 35%. In fact, 28% of the orthologs had alanine in this position, yet presumably can still form oligomers.

While many of the interfaces residues are conserved, the conservation of the actual contacts has not been significantly addressed. For example, OxyR has M230

interacting with L124. Do any orthologs have L230 interacting with M124? This would be particularly interesting to find residues that are involved in hydrogen bonding or salt-bridges. From the output of QConAlign, one would simply begin to count the number of ML and LM at the given position. Residues that are identical in all orthologs and involved in contacts would not need to be done, as their change is already known.

Several studies have emphasized the importance of structurally conserved positions in protein binding sites. The second question posed was "Are the same residues important in the oligomerization of a LTTR of unknown structure?" As discussed in Chapter III, we found that while most of the residues buried in the CynR interface correspond to equivalent residues in OxyR, there was no overlap in the residues important for oligomerization in OxyR and CynR., Why, then, do we not detect any overlap?

One possibility may lie in the protein that we used as a template. At the commencement of these studies, we were limited in crystal structures available to us. OxyR and CysB were the only ones available and for reasons we do not understand, CysB is not immune as a λcI repressor fusion. The response of OxyR to oxidative stress is characterized by the large rotation between the subunits. Now that we can examine the structure of CynR in the presence and absence of inducer, there is not a substantial difference between the residues in contact between the subunits, indicating that such rotation is not occurring in CynR. It would be interesting to know if other LTTRs that bind small molecules as their effectors shared hot spots with CynR. This would support the described hypothesize. The ten hot spots would be used to guide the initial

mutagenesis. Then, one could go after residues that are unique to the structures' contact profiles (as indicated by the shaded blue boxes in Figure 3.7.)

One advantage of alanine-scanning is that the contribution of the atoms beyond Cß can be evaluated. Yet, caution is warranted when evaluating the results of alanine-scanning, as illustrated by DeLano in 2002. As one is perturbing the interface, unexpected results can occur. For example, the mutation though located in the interface, can cause the protein to become unfolded. We suspect that this is the case in one of the Class I mutations in CynR, L228. This residue is located in the interface, but, examination of the atoms in both crystal forms shows that the atoms are completely buried in the monomer. Conversely, the mutation can reveal a negative design element that the protein has evolved (DeLano 2002). Thus, the mutation would cause appear to have no effect in oligomerization. If this type of interaction in OxyR was present, it might have been detected through the use of the regulatory domain in the lower expression vector, pLM1000, as wild-type OxyR is sensitive in this vector.

The repressor system is shown to be beneficial for the study of protein-protein interactions. In addition to identifying homotypic interactions in genome-wide screens (Marino-Ramirez and Hu 2002; Marino-Ramirez et al. 2004), it has been used extensively to study the oligomerization of the GCN4 leucine zippers, an elegant model for protein-protein interactions (Zeng et al. 1997; Zeng and Hu 1997). However, through the use of the cI repressor fusions, we were able to screen over 60 mutations and their contribution to oligomerization at different fusion protein concentrations. Evaluating this large number of mutations by traditional means would require

expression, purification and biochemical characterization of each individual mutation. This would be impractical for many LTTRs, which tend to be insoluble after over-expression. With the λcI repressor fusion, we were able to determine a qualitative strength of the mutations and focus the number of mutations for further studies. This is important in any system of proteins that are notoriously difficult to work with in vitro.

FUTURE DIRECTIONS

This project has laid the foundation for using the LTTRs as a model system for studying protein-protein interactions. While we have asked specific questions to address certain aspects of oligomerization in LTTRs, much more can still be done.

A long term goal of the oligomerization studies with OxyR was to utilize the negative-dominant system to study the residue's contribution to specificity. We did do pilot studies with the thioredoxin-OxyR fusions and λcI OxyR repressor fusions. However, these strains were repeatedly immune to the $\lambda i^{21}c$, something that was unexpected. The use of the regulatory domain as an inhibitor may serve as a better negative-dominant fusion. However, this result may be unavoidable with the thioredoxin-OxyR fusion, as the thioredoxin gene is a target of OxyR. Movement of the OxyR genes into the original cI^- vector may be necessary to continue negative dominant studies with the λcI repressor fusion (Zeng et al. 1997).

We also wanted to determine if the oligomerization changed with oxidative state. An *ahpCF* deletion lacks alkylhydroperoxide reductase and these strains have elevated endogenous levels of hydrogen peroxide (Rosner and Storz 1994). While we attempted

to perform the cross-streaks in a constitutively oxidizing environment, the control cI GCN4 repressor did not behave as expected. Additionally, we were limited by proteins that were only sensitive or immune in the vectors; that is, class 1a and class 3c. Therefore, moving the mutations out of the repressor system may be necessary for further characterization of the mutants. Giesela Storz's lab has used many different techniques to study OxyR *in vitro* and we have many of the strains and antibodies necessary in the lab, due to her generous gifts.

Parallel to studying the mutant OxyR's response to oxidative stress, we originally wanted to study CynR's response to its effector molecule, KCON. In the literature, we found many constructs that would allow for an *in vivo* negative-dominant system that was originally used to study CynR as a transcriptional regulator. Unfortunately, these constructs were lost in a freezer meltdown in the Fuch's lab at the University of Minnesota. One possibility we were pursuing with CynR was the use of a promoter-*gfp* fusions. We have obtained both the *cyn* promoter region fused to *gfp* in both directions. Expression of the mutants *in trans* should allow for assaying the mutant CynR's ability to behave as a transcriptional regulator in the presence and absence of KCON.

Finally, a long term goal is to take the information garnered from the alanine-scanning mutagenesis, as well as the resultant negative-dominant studies and reengineer the LTTRs to have alternate specificity. For example, thioredoxin CynR fusion is interacting with the λcI IciA fusion. It would be interesting to know whether hot spot mutations in CynR, as described in Chapter III, when made into a thioredoxin fusion, could disrupt this interaction. Conversely, do those mutations allow CynR to interact

with any other λcI LTTR repressor besides CynR and IciA? The same argument can be applied to the OxyR interactions.

REFERENCES

Anderson, P.M., Sung, Y.C., and Fuchs, J.A. 1990. The cyanase operon and cyanate metabolism. *FEMS Microbiol Rev* **7:** 247-252.

Aono, R., Tsukagoshi, N., and Yamamoto, M. 1998. Involvement of outer membrane protein TolC, a possible member of the *mar-sox* regulon, in maintenance and improvement of organic solvent tolerance of *Escherichia coli* K-12. *J Bacteriol* **180:** 938-944.

Appleman, J.R., Prendergast, N., Delcamp, T.J., Freisheim, J.H., and Blakley, R.L. 1988. Kinetics of the formation and isomerization of methotrexate complexes of recombinant human dihydrofolate reductase. *J Biol Chem* **263:** 10304-10313.

Argos, P. 1988. An investigation of protein subunit and domain interfaces. *Protein Eng* **2:** 101-113.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2:** 2006 0008.

Baccanari, D.P., Stone, D., and Kuyper, L. 1981. Effect of a single amino acid substitution on *Escherichia coli* dihydrofolate reductase catalysis and ligand binding. *J Biol Chem* **256:** 1738-1747.

Bahadur, R.P., Chakrabarti, P., Rodier, F., and Janin, J. 2003. Dissecting subunit interfaces in homodimeric proteins. *Proteins* **53:** 708-719.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* **28:** 235-242.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277:** 1453-1474.

Bogan, A.A., and Thorn, K.S. 1998. Anatomy of hot spots in protein interfaces. *J Mol Biol* **280:** 1-9.

Bremer, E., Silhavy, T.J., Weisemann, J.M., and Weinstock, G.M. 1984. Lambda placMu: a transposable derivative of bacteriophage lambda for creating *lacZ* protein fusions in a single step. *J Bacteriol* **158:** 1084-1093.

Bundy, B.M., Collier, L.S., Hoover, T.R., and Neidle, E.L. 2002. Synergistic transcriptional activation by one regulatory protein in response to two metabolites. *Proc Natl Acad Sci U S A* **99:** 7693-7698.

Bykowski, T., Ploeg, J.R.v.d., Iwanicka-Nowicka, R., and Hryniewicz, M.M. 2002. The switch from inorganic to organic sulphur assimilation in *Escherichia coli*: adenosine 5'-phosphosulphate (APS) as a signalling molecule for sulphate excess. *Molecular Microbiology* **43:** 1347-1358.

Carter, E.L., Jager, L., Gardner, L., Hall, C.C., Willis, S., and Green, J.M. 2007. *Escherichia coli abg* genes enable uptake and cleavage of the folate catabolite p-aminobenzoyl-glutamate. *J Bacteriol* **189:** 3329-3334.

Cathelyn, J.S., Crosby, S.D., Lathem, W.W., Goldman, W.E., and Miller, V.L. 2006. RovA, a global regulator of *Yersinia pestis*, specifically required for bubonic plague. *Proc Natl Acad Sci U S A* **103:** 13514-13519.

Celis, R.T. 1999. Repression and activation of arginine transport genes in *Escherichia col*i K 12 by the ArgP protein. *J Mol Biol* **294:** 1087-1095.

Chakrabarti, P., and Janin, J. 2002. Dissecting protein-protein recognition sites. *Proteins* **47:** 334-343.

Chattopadhyay, S., Wu, Y., and Datta, P. 1997. Involvement of Fnr and ArcA in anaerobic expression of the *tdc* operon of *Escherichia coli*. *J Bacteriol* **179:** 4868-4873.

Choi, H., Kim, S., Mukhopadhyay, P., Cho, S., Woo, J., Storz, G., and Ryu, S. 2001. Structural basis of the redox switch in the OxyR transcription factor. *Cell* **105:** 103-113.

Chothia, C. 1974. Hydrophobic bonding and accessible surface area in proteins. *Nature* **248:** 338-339.

Chothia, C., and Janin, J. 1975. Principles of protein-protein recognition. *Nature* **256:** 705-708.

Christman, M.F., Storz, G., and Ames, B.N. 1989. OxyR, a positive regulator of hydrogen peroxide-inducible genes in *Escherichia coli* and *Salmonella typhimurium*, is homologous to a family of bacterial regulatory proteins. *Proc Natl Acad Sci U S A* **86:** 3484-3488.

Clackson, T., and Wells, J.A. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science* **267:** 383-386.

Colyer, T.E., and Kredich, N.M. 1994. Residue threonine-149 of the *Salmonella typhimurium* CysB transcription activator: mutations causing constitutive expression of positively regulated genes of the cysteine regulon. *Mol Microbiol* **13:** 797-805.

Cusa, E., Obradors, N., Baldoma, L., Badia, J., and Aguilar, J. 1999. Genetic analysis of a chromosomal region containing genes required for assimilation of allantoin nitrogen and linked glyoxylate metabolism in *Escherichia coli. J Bacteriol* **181:** 7479-7484.

DeLano, W.L. 2002. Unraveling hot spots in binding interfaces: progress and challenges. *Current Opinion in Structural Biology* **12:** 14-20.

Dougherty, T.J., and Pucci, M.J. 1994. Penicillin-binding proteins are regulated by *rpoS* during transitions in growth states of *Escherichia coli. Antimicrob Agents Chemother* **38:** 205-210.

Ezezika, O.C., Haddad, S., Clark, T.J., Neidle, E.L., and Momany, C. 2007. Distinct effector-binding sites enable synergistic transcriptional activation by BenM, a LysR-type regulator. *J Mol Biol* **367:** 616-629.

Fang, M., Majumder, A., Tsai, K.J., and Wu, H.Y. 2000. ppGpp-dependent *leuO* expression in bacteria under stress. *Biochem Biophys Res Commun* **276:** 64-70.

Fischer, T.B., Arunachalam, K.V., Bailey, D., Mangual, V., Bakhru, S., Russo, R., Huang, D., Paczkowski, M., Lalchandani, V., Ramachandra, C., et al. 2003. The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces. *Bioinformatics* **19:** 1453-1454.

Fischer, T.B., Holmes, J.B., Miller, I.R., Parsons, J.R., Tung, L., Hu, J.C., and Tsai, J. 2006. Assessing methods for identifying pair-wise atomic contacts across binding interfaces. *J Struct Biol* **153:** 103-112.

Fralick, J.A. 1996. Evidence that TolC is required for functioning of the Mar/AcrAB efflux pump of *Escherichia coli*. *J Bacteriol* **178:** 5803-5805.

Ganduri, Y.L., Sadda, S.R., Datta, M.W., Jambukeswaran, R.K., and Datta, P. 1993. TdcA, a transcriptional activator of the *tdcABC* operon of *Escherichia coli*, is a member of the LysR family of proteins. *Mol Gen Genet* **240:** 395-402.

Ghrist, A.C., Heil, G., and Stauffer, G.V. 2001. GcvR interacts with GcvA to inhibit activation of the *Escherichia coli* glycine cleavage operon. *Microbiology* **147:** 2215-2221.

Goller, C., Wang, X., Itoh, Y., and Romeo, T. 2006. The cation-responsive protein NhaR of *Escherichia coli* activates *pgaABCD* transcription required for production of the biofilm adhesin poly-{beta}-1,6-N-acetyl-D-glucosamine. *J. Bacteriol.* **188:** 8022-8032.

Guharoy, M., and Chakrabarti, P. 2005. Conservation and relative importance of residues across protein-protein interfaces. *Proc Natl Acad Sci U S A* **102:** 15447-15452.

Hagewood, B.T., Ganduri, Y.L., and Datta, P. 1994. Functional analysis of the *tdcABC* promoter of *Escherichia col*i: roles of TdcA and TdcR. *J. Bacteriol.* **176:** 6214-6220.

Heil, G., Stauffer, L.T., and Stauffer, G.V. 2002. Glycine binds the transcriptional accessory protein GcvR to disrupt a GcvA/GcvR interaction and allow GcvA-mediated activation of the *Escherichia coli gcvTHP* operon. *Microbiology* **148:** 2203-2214.

Henikoff, S., Haughn, G.W., Calvo, J.M., and Wallace, J.C. 1988. A large family of bacterial activator proteins. *Proc Natl Acad Sci U S A* **85:** 6602-6606.

Henrick, K., and Thornton, J.M. 1998. PQS: a protein quaternary structure file server. *Trends Biochem Sci* **23:** 358-361.

Howell, E.E., Foster, P.G., and Foster, L.M. 1988. Construction of a dihydrofolate reductase-deficient mutant of *Escherichia coli* by gene replacement. *J Bacteriol* **170:** 3040-3045.

Hryniewicz, M.M., and Kredich, N.M. 1994. Stoichiometry of binding of CysB to the *cysJIH, cysK*, and *cysP* promoter regions of *Salmonella typhimurium*. *J Bacteriol* **176:** 3673-3682.

Hu, J.C., Newell, N.E., Tidor, B., and Sauer, R.T. 1993. Probing the roles of residues at the **e** and **g** positions of the GCN4 leucine zipper by combinatorial mutagenesis. *Protein Sci* **2:** 1072-1084.

Hu, J.C., O'Shea, E.K., Kim, P.S., and Sauer, R.T. 1990. Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science* **250:** 1400-1403.

Hu, Z., Ma, B., Wolfson, H., and Nussinov, R. 2000. Conservation of polar residues as hot spots at protein interfaces. *Proteins* **39:** 331-342.

Hussein, M.J., Green, J.M., and Nichols, B.P. 1998. Characterization of mutations that allow p-aminobenzoyl-glutamate utilization by *Escherichia coli*. *J. Bacteriol.* **180:** 6260-6268.

Hwang, D.S., Thony, B., and Kornberg, A. 1992. IciA protein, a specific inhibitor of initiation of *Escherichia coli* chromosomal replication. *J Biol Chem* **267:** 2209-2213.

Izumiya, H., Terajima, J., Wada, A., Inagaki, Y., Itoh, K.I., Tamura, K., and Watanabe, H. 1997. Molecular typing of enterohemorrhagic *Escherichia coli* O157:H7 isolates in Japan by using pulsed-field gel electrophoresis. *J Clin Microbiol* **35:** 1675-1680.

Jagura-Burdzy, G., and Hulanicka, D. 1981. Use of gene fusions to study expression of *cysB*, the regulatory gene of the cysteine regulon. *J Bacteriol* **147:** 744-751.

Janin, J., Miller, S., and Chothia, C. 1988. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol* **204:** 155-164.

Jones, S., and Thornton, J.M. 1995. Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol* **63:** 31-65.

Jones, S., and Thornton, J.M. 1996. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **93:** 13-20.

Jones, S., and Thornton, J.M. 2000. Analysis and clssification or protein-protein interactions from a structural perspective. In *Protein-protein Recognition*. (ed. C. Kleanthous), pp. 33-59. Oxford University Press, Oxford.

Jorgensen, C., and Dandanell, G. 1999. Isolation and characterization of mutations in the *Escherichia coli* regulatory protein XapR. *J Bacteriol* **181:** 4397-4403.

Jourdan, A.D., and Stauffer, G.V. 1998. Mutational analysis of the transcriptional regulator GcvA: amino acids important for activation, repression, and DNA binding. *J Bacteriol* **180:** 4865-4871.

Jourdan, A.D., and Stauffer, G.V. 1999. GcvA-mediated activation of *gcvT-lacZ* expression involves the carboxy-terminal domain of the alpha subunit of RNA polymerase. *FEMS Microbiol Lett* **181:** 307-312.

Keskin, O., Ma, B., and Nussinov, R. 2005. Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol* **345:** 1281-1294.

Klauck, E., Bohringer, J., and Hengge-Aronis, R. 1997. The LysR-like regulator LeuO in *Escherichia coli* is involved in the translational regulation of *rpoS* by affecting the expression of the small regulatory DsrA-RNA. *Mol Microbiol* **25:** 559-569.

Klotz, I.M., Darnall, D.W., and Langerman, N.R. 1975. Quaternary structure of proteins. In *The Proteins*, 3d ed. (ed. H.a.H. Neurath, Robert), pp. 293-411. Academic Press, New York.

Kovacikova, G., and Skorupski, K. 1999. A *Vibrio cholerae* LysR homolog, AphB, cooperates with AphA at the *tcpPH* promoter to activate expression of the ToxR virulence cascade. *J. Bacteriol.* **181:** 4250-4256.

Kredich, N.M. 1992. The molecular basis for positive regulation of *cys* promoters in *Salmonella typhimurium* and *Escherichia coli*. *Mol Microbiol* **6:** 2747-2753.

Kullik, I., Stevens, J., Toledano, M.B., and Storz, G. 1995a. Mutational analysis of the redox-sensitive transcriptional regulator OxyR: regions important for DNA binding and multimerization. *J Bacteriol* **177:** 1285-1291.

Kullik, I., Toledano, M.B., Tartaglia, L.A., and Storz, G. 1995b. Mutational analysis of the redox-sensitive transcriptional regulator OxyR: regions important for oxidation and transcriptional activation. *J Bacteriol* **177:** 1275-1284.

Lamblin, A.F., and Fuchs, J.A. 1994. Functional analysis of the *Escherichia coli* K-12 *cyn* operon transcriptional regulation. *J Bacteriol* **176:** 6613-6622.

Lee, Y., Lee, H., Yim, J., and Hwang, D. 1997. The binding of two dimers of IciA protein to the *dnaA* promoter 1P element enhances the binding of RNA polymerase to the *dnaA* promoter 1P. *Nucleic Acids Res* **25:** 3486-3489.

Lee, Y.S., Kim, H., and Hwang, D.S. 1996. Transcriptional activation of the *dnaA* gene encoding the initiator for *oriC* replication by IciA protein, an inhibitor of *in vitro* oriC replication in *Escherichia coli. Mol Microbiol* **19:** 389-396.

Levy, S.B. 1992. Active efflux mechanisms for antimicrobial resistance. *Antimicrob. Agents Chemother.* **36:** 695-703.

Lewis, K. 1994. Multidrug resistance pumps in bacteria: variations on a theme. *Trends Biochem Sci* **19:** 119-123.

Liu, Y., and Eisenberg, D. 2002. 3D domain swapping: as domains continue to swap. *Protein Sci* **11:** 1285-1299.

Lochowska, A., Iwanicka-Nowicka, R., Plochocka, D., and Hryniewicz, M.M. 2001. Functional dissection of the LysR-type CysB transcriptional regulator. Regions

important for DNA binding, inducer response, oligomerization, and positive control. *J Biol Chem* **276:** 2098-2107.

Lomovskaya, O., Lewis, K., and Matin, A. 1995. EmrR is a negative regulator of the *Escherichia coli* multidrug resistance pump EmrAB. *J Bacteriol* **177:** 2328-2334.

Lorenz, E., and Stauffer, G.V. 1995. Characterization of the MetR binding sites for the *glyA* gene of *Escherichia coli*. *J Bacteriol* **177:** 4113-4120.

Ma, D., Cook, D.N., Alberti, M., Pon, N.G., Nikaido, H., and Hearst, J.E. 1993. Molecular cloning and characterization of *acrA* and *acrE* genes of *Escherichia coli*. *J Bacteriol* **175:** 6299-6313.

Mares, R., Urbanowski, M.L., and Stauffer, G.V. 1992. Regulation of the *Salmonella typhimurium metA* gene by the metR protein and homocysteine. *J Bacteriol* **174:** 390-397.

Marino-Ramirez, L., Campbell, L., and Hu, J.C. 2003. Screening peptide/protein libraries fused to the lambda repressor DNA-binding domain in *E. coli* cells. *Methods Mol Biol* **205:** 235-250.

Marino-Ramirez, L., and Hu, J.C. 2002. Isolation and mapping of self-assembling protein domains encoded by the *Saccharomyces cerevisiae* genome using lambda repressor fusions. *Yeast* **19:** 641-650.

Marino-Ramirez, L., Minor, J.L., Reading, N., and Hu, J.C. 2004. Identification and mapping of self-assembling protein domains encoded by the *Escherichia coli* K-12 genome by use of lambda repressor fusions. *J Bacteriol* **186:** 1311-1319.

Milkman, R. 1973. Electrophoretic variation in *Escherichia coli* from natural sources. *Science* **182:** 1024-1026.

Miller, J.H. 1972. *Experiments in molecular genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Morona, R., Manning, P.A., and Reeves, P. 1983. Identification and characterization of the TolC protein, an outer membrane protein from *Escherichia coli*. *J Bacteriol* **153:** 693-699.

Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K., and Senda, T. 2003. Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *J Mol Biol* **328:** 555-566.

Muse, W.B., and Bender, R.A. 1998. The *nac* (nitrogen assimilation control) gene from *Escherichia coli*. *J Bacteriol* **180:** 1166-1173.

Muse, W.B., Rosario, C.J., and Bender, R.A. 2003. Nitrogen regulation of the *codBA* (cytosine deaminase) operon from *Escherichia coli* by the nitrogen assimilation control protein, NAC. *J Bacteriol* **185:** 2920-2926.

Nagel de Zwaig, R., and Luria, S.E. 1967. Genetics and physiology of colicin-tolerant mutants of *Escherichia coli*. *J. Bacteriol.* **94:** 1112-1123.

Nikaido, H. 1994. Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science* **264:** 382-388.

Nikaido, H. 1996. Multidrug efflux pumps of gram-negative bacteria. *J Bacteriol* **178:** 5853-5859.

Nikaido, H., and Vaara, M. 1985. Molecular basis of bacterial outer membrane permeability. *Microbiol Rev* **49:** 1-32.

Padan, E., Maisler, N., Taglicht, D., Karpel, R., and Schuldiner, S. 1989. Deletion of *ant* in *Escherichia coli* reveals its function in adaptation to high salinity and an alternative Na+/H+ antiporter system(s). *J Biol Chem* **264:** 20297-20302.

Perkins, J.D., Heath, J.D., Sharma, B.R., and Weinstock, G.M. 1993. XbaI and BlnI genomic cleavage maps of *Escherichia coli* K-12 strain MG1655 and comparative analysis of other strains. *J Mol Biol* **232:** 419-445.

Plamann, M.D., and Stauffer, G.V. 1989. Regulation of the *Escherichia coli glyA* gene by the *metR* gene product and homocysteine. *J Bacteriol* **171:** 4958-4962.

Poggio, S., Domeinzain, C., Osorio, A., and Camarena, L. 2002. The nitrogen assimilation control (Nac) protein represses *asnC* and *asnA* transcription in *Escherichia coli*. *FEMS Microbiol Lett* **206:** 151-156.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res* **33:** W116-120.

Rahav-Manor, O., Carmel, O., Karpel, R., Taglicht, D., Glaser, G., Schuldiner, S., and Padan, E. 1992. NhaR, a protein homologous to a family of bacterial regulatory proteins (LysR), regulates *nhaA*, the sodium proton antiporter gene in *Escherichia coli*. *J Biol Chem* **267:** 10433-10438.

Rhee, K.Y., Senear, D.F., and Hatfield, G.W. 1998. Activation of gene expression by a ligand-induced conformational change of a protein-DNA complex. *J Biol Chem* **273:** 11257-11266.

Rintoul, M.R., Cusa, E., Baldoma, L., Badia, J., Reitzer, L., and Aguilar, J. 2002. Regulation of the *Escherichia coli* allantoin regulon: coordinated function of the repressor AllR and the activator AllS. *J Mol Biol* **324:** 599-610.

Rosario, C.J., and Bender, R.A. 2005. Importance of tetramer formation by the nitrogen assimilation control protein for strong repression of glutamate dehydrogenase formation in *Klebsiella pneumoniae*. *J. Bacteriol.* **187:** 8291-8299.

Rosner, J.L., and Storz, G. 1994. Effects of peroxides on susceptibilities of *Escherichia coli* and *Mycobacterium smegmatis* to isoniazid. *Antimicrob Agents Chemother* **38:** 1829-1833.

Sawers, G. 2001. A novel mechanism controls anaerobic and catabolite regulation of the *Escherichia coli tdc* operon. *Mol Microbiol* **39:** 1285-1298.

Schell, M.A. 1993. Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol* **47:** 597-626.

Schell, M.A., Brown, P.H., and Raju, S. 1990. Use of saturation mutagenesis to localize probable functional domains in the NahR protein, a LysR-type transcription activator. *J Biol Chem* **265:** 3844-3850.

Seeger, C., Poulsen, C., and Dandanell, G. 1995. Identification and characterization of genes (*xapA, xapB*, and *xapR*) involved in xanthosine catabolism in *Escherichia coli*. *J Bacteriol* **177:** 5506-5516.

Selander, R.K., and Levin, B.R. 1980. Genetic diversity and structure in *Escherichia coli* populations. *Science* **210:** 545-547.

Smirnova, I.A., Dian, C., Leonard, G.A., McSweeney, S., Birse, D., and Brzezinski, P. 2004. Development of a bacterial biosensor for nitrotoluenes: the crystal structure of the transcriptional regulator DntR. *J Mol Biol* **340:** 405-418.

Stec, E., Witkowska-Zimny, M., Hryniewicz, M.M., Neumann, P., Wilkinson, A.J., Brzozowski, A.M., Verma, C.S., Zaim, J., Wysocki, S., and Bujacz, G.D. 2006. Structural basis of the sulphate starvation response in *E. coli*: crystal structure and mutational analysis of the cofactor-binding domain of the Cbl transcriptional regulator. *J Mol Biol* **364:** 309-322.

Stragier, P., and Patte, J.C. 1983. Regulation of diaminopimelate decarboxylase synthesis in *Escherichia coli.* III. Nucleotide sequence and regulation of the *lysR* gene. *J Mol Biol* **168:** 333-350.

Stragier, P., Richaud, F., Borne, F., and Patte, J.C. 1983. Regulation of diaminopimelate decarboxylase synthesis in *Escherichia coli.* I. Identification of a *lysR* gene encoding an activator of the *lysA* gene. *J Mol Biol* **168:** 307-320.

Sung, Y.C., and Fuchs, J.A. 1988. Characterization of the *cyn* operon in *Escherichia coli* K12. *J Biol Chem* **263:** 14769-14775.

Sung, Y.C., and Fuchs, J.A. 1992. The *Escherichia coli* K-12 *cyn* operon is positively regulated by a member of the *lysR* family. *J Bacteriol* **174:** 3645-3650.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28:** 33-36.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. *Science* **278:** 631-637.

Tatusov, R.L., Natale, D.A., Garkavt sev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29:** 22-28.

Thony, B., Hwang, D.S., Fradkin, L., and Kornberg, A. 1991. *iciA*, an *Escherichia coli* gene encoding a specific inhibitor of chromosomal initiation of replication *in vitro*. *Proc Natl Acad Sci U S A* **88:** 4066-4070.

Tsai, C.J., Lin, S.L., Wolfson, H.J., and Nussinov, R. 1996. Protein-protein interfaces: architectures and interactions in protein-protein interfaces and in protein cores. Their similarities and differences. *Crit Rev Biochem Mol Biol* **31:** 127-152.

Tsai, J. 2008. Binding Interface Database Wiki. http://tsailab.org/wikiBID/index.php/Main_Page.

Turlin, E., Sismeiro, O., Le Caer, J.P., Labas, V., Danchin, A., and Biville, F. 2005. 3-phenylpropionate catabolism and the *Escherichia col*i oxidative stress response. *Res Microbiol* **156:** 312-321.

Tyrrell, R., Verschueren, K.H., Dodson, E.J., Murshudov, G.N., Addy, C., and Wilkinson, A.J. 1997. The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Structure* **5:** 1017-1032.

Vaara, M. 1993. Antibiotic-supersusceptible mutants of *Escherichia coli* and *Salmonella typhimurium*. *Antimicrob Agents Chemother* **37:** 2255-2260.

van Der Ploeg, J.R., Iwanicka-Nowicka, R., Bykowski, T., Hryniewicz, M.M., and Leisinger, T. 1999. The *Escherichia coli ssuEADCB* gene cluster is required for the utilization of sulfur from aliphatic sulfonates and is regulated by the transcriptional activator Cbl. *J Biol Chem* **274:** 29358-29365.

Van Dyk, T.K., Templeton, L.J., Cantera, K.A., Sharpe, P.L., and Sariaslani, F.S. 2004. Characterization of the *Escherichia coli* AaeAB efflux pump: a metabolic relief valve? *J Bacteriol* **186:** 7196-7204.

Vershon, A.K., Bowie, J.U., Karplus, T.M., and Sauer, R.T. 1986. Isolation and analysis of arc repressor mutants: evidence for an unusual mechanism of DNA binding. *Proteins* **1:** 302-311.

Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. 2000. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* **328:** 575-592.

Wandersman, C., and Delepelaire, P. 1990. TolC, an *Escherichia coli* outer membrane protein required for hemolysin secretion. *Proc Natl Acad Sci U S A* **87:** 4776-4780.

Wang, X., Mukhopadhyay, P., Wood, M.J., Outten, F.W., Opdyke, J.A., and Storz, G. 2006. Mutational analysis to define an activating region on the redox-sensitive transcriptional regulator OxyR. *J Bacteriol* **188:** 8335-8342.

Whittam, T.S., Ochman, H., and Selander, R.K. 1983. Multilocus genetic structure in natural populations of *Escherichia coli*. *Proc Natl Acad Sci U S A* **80:** 1751-1755.

Wilson, R.L., Urbanowski, M.L., and Stauffer, G.V. 1995. DNA binding sites of the LysR-type regulator GcvA in the *gcv* and *gcvA* control regions of *Escherichia coli*. *J Bacteriol* **177:** 4940-4946.

Wodak, S.J., and Janin, J. 2002. Structural basis of macromolecular recognition. *Adv Protein Chem* **61:** 9-73.

Xiao, G., Deziel, E., He, J., Lepine, F., Lesic, B., Castonguay, M.H., Milot, S., Tampakaki, A.P., Stachel, S.E., and Rahme, L.G. 2006a. MvfR, a key *Pseudomonas aeruginosa* pathogenicity LTTR-class regulatory protein, has dual ligands. *Mol Microbiol* **62:** 1689-1699.

Xiao, G., He, J., and Rahme, L.G. 2006b. Mutation analysis of the *Pseudomonas aeruginosa mvfR* and *pqsABCDE* gene promoters demonstrates complex quorum-sensing circuitry. *Microbiology* **152:** 1679-1686.

Zeng, X., Herndon, A.M., and Hu, J.C. 1997. Buried asparagines determine the dimerization specificities of leucine zipper mutants. *Proc Natl Acad Sci U S A* **94:** 3673-3678.

Zeng, X., and Hu, J.C. 1997. Detection of tetramerization domains *in vivo* by cooperative DNA binding to tandem lambda operator sites. *Gene* **185:** 245-249.

Zheng, M., Aslund, F., and Storz, G. 1998. Activation of the OxyR transcription factor by reversible disulfide bond formation. *Science* **279:** 1718-1721.

Zimmer, D.P., Soupene, E., Lee, H.L., Wendisch, V.F., Khodursky, A.B., Peter, B.J., Bender, R.A., and Kustu, S. 2000. Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. *Proc Natl Acad Sci U S A* **97:** 14674-14679.

APPENDIX A

PERL SCRIPTS WRITTEN FOR THIS PROJECT

QConAlign.pl

```
#!/usr/bin/perl


# Qcons alignv1.pl
#
# This is a program to map the contact profile from a member
# of a COG to other members of the COG for which the structure
# is unavailable.
# The sequence alignment is based on CLUSTALW, for now.

# ppialignv2.pl has been modified such that the numbers
generated and output
# are based on the universal coordinate system.  Further, all of
the translated
# output is put into three files.

# ppialinv3.pl has been modified from ppialignv2.pl to account
for the fact that
# chains may not be JUST A or B. In this case, it was chain P
that was needed.

#Qconalign.pl has been evolved to utilize the output from Qcons.
It has the same
#functionality as ppialign, just takes into account the
different output format of Qcons.
# gsk 07.20.04

#Usage:
#perl Qconalign.pl < [alignment file] t [protein template] res
[Qcon residue file]

# ---------------------------------------------------------
# ---------------- Input Variables ----------------------
# ---------------------------------------------------------

# if pdb file residue numbering or register is not the same as
in the sequence file
# reset here
$pdboffset = varafteroption("o", 1, "sequence offset of the pdb
file to sequence file ");
```

```
# debugging
$debug = 0;

# Asks which sequence in the alignment is the template sequence.
#printf STDERR "yo\n";

$template = varafteroption("t", "none", "template defined as ");
if ($template eq "none") {
    printf "Jim says he needs a template name after option
t!\n";
    die;
}

# Defining names of Qcon files (for input)
# voronoi interactions
$byres = varafteroption("res", "CbnR.res", "The file being used
is ");
if (! -r $byres) {
    printf "Cannnot read $byres!\n";
    die;
}




# ----------------------------------------------------------------
# ----Read in Sequence alignment from STDIN --------------------
# ----------------------------------------------------------------

# stores COG multiples sequence alignment information in
variable seq
# initializes hash to zero

%seq = ();

# takes input from  STDIN
# removes header and white space and joins the sequences
together.
while (<STDIN>) {

   # printf "Hi\n";
    next if /^CLUSTAL /;
    chomp;
    @tmp = split ' ';
    next unless scalar(@tmp) == 2;
    $seqp=sprintf ("%s%s", $seq{$tmp[0]},$tmp[1]);
}

# This is checking to see if template name is in the file
```

```
$isnotthere=1;
foreach $a (keys %seq) {
    if ($a eq $template) {
        $isnotthere=0;
     last;
    }
}
if($isnotthere) { die "didn't have $template in the file!
exiting ...\n"; }

# Creating master index to template sequence
# note $template depends on above
@seqarray= split '', $seq{$template};
$sequenceindex=$pdboffset; @masterindex=();
for ($resnum=0;$resnum<@seqarray;$resnum++) {
    $resname=$seqarray[$resnum];
    if ($resname ne "-") {
        $masterindex[$sequenceindex]=$resnum;
     $masterseq[$sequenceindex]=$resname;
        $sequenceindex++;
    }
}


# ------------------------------
# READ IN DATA FROM ppiContacts
# ------------------------------

# Initializing variables
%resnameA = (); %resnumA = ();
%resnameB = (); %resnumB = ();
%chainA = (); %chainO = ();




# Commands for calling subroutines (below)


readbyres();


# this is just a check
if ($debug) {
    foreach $a (sort ascending keys %resnameA) {
     printf STDERR "$a      A $resnumA{$a} $resnameA{$a} to B
$resnumB{$a} $resnameB{$a}\n";
```

```
     printf STDERR "$a      A $masterseq[$resnumA{$a}]
$masterindex[$resnumA{$a}]  to B  $masterseq[$resnumB{$a}]
$masterindex[$resnumB{$a}]\n";
    }
}

# ------------------------------------------------------------
# going over the multiple sequence alignment & doing replacement
# ------------------------------------------------------------

open(OO,">${template}.resvor") || die "couldn't open
${template}.resvor\n";
printf OO "Protein Name, INnum, Chain, ResNum, MIN, Res to
INnum, Chain, ResNum, MIN, Res  \n";


foreach $protein (keys %seq) {
    #next if $protein eq $template;
    $sequences=$seq{$protein};
    printf STDERR "$protein $sequences\n";
    @seqpos = split //, $sequences;
    @seqposnum = (); $nn=0;

    for ($ii=0,$nn=0;$ii<@seqpos;$ii++) {
     $residue = $seqpos[$ii];
     if ($residue eq '-') {
         $seqposnum[$ii] = $nn + 0.5;
     }
     else {
         $nn++;
         $seqposnum[$ii]= $nn;
     }
    }



#outputting the information for the hSA
#This will output the protein name, index number, the chain
name, the original residue number, the masterindex number, the
residue at that position in the protein to
#the chain name, the original residue number, the master index
number and the residue at that position in the protein.

foreach $a (sort ascending keys %resnameA) {
     #printf STDERR
"$a\tA\t$masterindex[$resnumA{$a}]\t$resnameA{$a}\tto\tB\t$maste
rindex[$resnumB{$a}]\t$resnameB{$a}\n";
     #printf STDERR
"$a\tA\t$masterindex[$resnumA{$a}]\t$seqpos[$masterindex[$resnum
```

```
A{$a}]]\tto\tB\t$masterindex[$resnumB{$a}]\t$seqpos[$masterindex
[$resnumB{$a}]]\n";

     printf OO "%15s %15d  %6s %6s  %7d  %4s  to  %6s %6s %7d
%4s\n", $protein, $a, $chainA{$a}, $resnumA{$a},
$masterindex[$resnumA{$a}], $seqpos[$masterindex[$resnumA{$a}]],
$chainO{$a}, $resnumB{$a}, $masterindex[$resnumB{$a}],
$seqpos[$masterindex[$resnumB{$a}]];
    }



}
close (OO);



# -------- SUBROUTINES -----------

# subroutine to read in hSA
#     0     1  2   3 4  5   6   7   8
#     4587  10 LEU A -  347 THR B 100.00
sub readbyres {
    printf STDERR "byres\n";
    open (SA, $byres);
    while(<SA>) {
        chomp;
        next if $_ =~ /^\#/;
           #next unless $_ =~ / \- /;

        my @column = split(' ', $_);

           my $identnum = $column[0];
        my $firstchain = $column[3];

# fill data for chain A
        if($firstchain eq 'A') {
            $resnumA{$identnum} = $column[1];
            $resnameA{$identnum} = $column[2];
                $chainA{$identnum}= $column[3];

            $resnumB{$identnum} = $column[5];
            $resnameB{$identnum} = $column[6];
                $chainO{$identnum} = $column[7];

        } else {
            $resnumB{$identnum} = $column[1];
            $resnameB{$identnum} = $column[2];
               $chainO{$identnum} = $column[3];
```

```
            $resnumA{$identnum} = $column[5];
               $resnameA{$identnum} = $column[6];
                $chainA{$identnum} = $column[7];
        }
    }
    close(SA);
}


# subroutine to return variable after command line argument
# usage: $stt = varafteroption("stt", 1.00, "start oxygen radii
at ");

sub varafteroption {
    my $changed=0;
    for($ii=0;$ii<@ARGV;$ii++)  {
        if($ARGV[$ii] eq $_[0])  {
            $var = $ARGV[$ii+1]; $changed++;
            print STDERR "NEW $_[2]($_[0]) $var\n";
            last;
        }
    }

    if(!$changed)  {
        $var = $_[1];
        print STDERR "DEFAULT $_[2]($_[0]) $var\n";
    }
    return $var;
}

sub ascending { $a <=> $b; }
```

QconAAtally.pl

```
#!/usr/bin/perl

#QconAAtally.pl
#April 10, 2006

#usage perl QconAAtally.pl < (Alignment file)

#output: Column aa and #, aa and #, aa and # etc.


open(OO,">AAcount.txt") || die "couldn't open AAcount.txt!:
$!\n";
# printf OO "Hi";

# takes input from  STDIN
# removes header and white space and joins the sequences
together.
# reads data into hash called $seq{}

%seq = ();
while (<STDIN>) {
    next if /^CLUSTAL /;

#skip blank lines
    next if /^(\s)*$/;

    # printf stderr "%s", $_;
    chomp;
    @tmp = split ' ';
    next unless scalar(@tmp) == 2;
    $seq{$tmp[0]}= sprintf ("%s%s", $seq{$tmp[0]},$tmp[1]);
}

#just testing to see if it's actually joining....
#It is actually working....

# foreach $a (keys %seq) {
#     printf OO "%30s %s\n", $a, $seq{$a};
# }

# $msa[$position]{$aa} for reading array
# but from where does msa come?

#The -1 is saying if this is the fist time you've seen this.
```

```
$msa = ();
$seqlength = -1;

foreach $ss (keys %seq) {
#spliting the sequence into individual characters
    @tmp = split //, $seq{$ss};

    # below is a safety check for sequence length
    if ($seqlength == -1) {
      $seqlength = scalar(@tmp);
    }

    elsif ($seqlength != scalar(@tmp)) {
      printf stderr "warning %s sequence %d is a different length
than %d\n", $ss, scalar(@tmp), $seqlength;
    }

    # for ($position=0; $position<@tmp; $tmp) {   this was
wrong, needed to increment position
    for ($position=0;$position<scalar(@tmp);$position++) {
      $aa=$tmp[$position];                  # getting amino acid
type
      $msa[$position]{$aa}++;               # incrementing that
type
    }
}


# Dereferencing
# It used to be $position+1 so that it started at 1; however
this made the
# MIN be off one from QconAlign MIN.  Therefore, the positioning
will start at 0.

for ($position=0; $position<scalar(@msa); $position++) {
    printf OO "%3d ", $position;
    foreach $aa (keys %{@msa[$position]}) {
      printf OO "| %s %-2d ", $aa, $msa[$position]{$aa};
    }

    printf OO "\n";
}

close OO;
```

APPENDIX B

OxyR ORTHOLOGS

>EOxyR
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGMLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVPMADLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPRRTIGLVYRPGSPLRS
RYEQLAEAIRARMDGHFDKVLKQAV

>gi|83569771|ref|ZP_00921220.1| COG0583: Transcriptional regulator
[Shigella dysenteriae 1012]
MNYRGNGGWIMNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQA
GMLLVDQARTVLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTH
QLLAQLDSGKLDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVPMADLAGEKLLMLEDGHCL
RDQAMGFCFEAGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPRRTIGL
VYRPGSPLRSRYEQLAEAIRARMDGHFDKVLKQAV

>gi|75176648|ref|ZP_00696776.1| COG0583: Transcriptional regulator
[Shigella boydii BS512]
MNYRGDGGWIMNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQA
GMLLVDQARTVLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTH
QLLAQLDSGKLDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVPMADLAGEKLLMLEDGHCL
RDQAMGFCFEAGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPRRTIGL
VYRPGSPLRSRYEQLAEAIRARMDGHFDKVLKQAV

>gi|26250736|ref|NP_756776.1| Hydrogen peroxide-inducible genes
activator [Escherichia coli CFT073]
MLPIAANYRGNGGWIMNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKV
LFTQAGMLLVDQARTVLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLH
EAQTHQLLAQLDSGKLDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVPMADLAGEKLLMLE
DGHCLRDQAMGFCFEAGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPR
RTIGLVYRPGSPLRSRYEQLAEAIRARMDGHFDKVLKQAV

>gi|75238409|ref|ZP_00722408.1| COG0583: Transcriptional regulator
[Escherichia coli E110019]
MNYRGDGGWIMNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQA
GMLLVDQARTVLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTH
QLLAQLDSGKLDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVPMADLAGEKLLMLEDGHCL
RDQAMGFCFEAGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPRRTIGL
VYRPGSPLRSRYEQLAEAIRERMDGHFDKVLKQAV

>gi|15804557|ref|NP_290598.1| activator, hydrogen peroxide-inducible
genes [Escherichia coli O157:H7 EDL933]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGMLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVPMADLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPRRTIGLVYRPGSPLRS
RYEQLAEAIRARMDGHFDKVLKQAV

>gi|74314463|ref|YP_312882.1| activator, hydrogen peroxide-inducible genes [Shigella sonnei Ss046]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGMLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVPMADLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPRRTIGLVYRPGSPLRS
RYEQLAEAIRARMDGYFDKVLKQAV

>gi|147042|gb|AAA24257.1| oxyR protein
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGMLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCVILALVKESERFIEVPLFDEPMLLAIYEDHPWANRECVPMADLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPRRTIGLVYRPGSPLRS
RYEQLAEAIRARMDGHFDKVLKQAV

>gi|146880|gb|AAA24176.1| morphology control protein (put.); putative
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGMLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVPMADLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPEAKRDGVVYLPCIKPEPRRTIGLVYRPGSPLRS
RYEQLAEAIRARMDGHFDKVLKQAV

>gi|110807820|ref|YP_691340.1| OxyR [Shigella flexneri 5 str. 8401]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGMLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVLMADLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPRRTIGLVYRPGSPLRS
RYEQLAEAIRARMDGHFDKVLKQAV

>gi|24115255|ref|NP_709765.1| activator of hydrogen peroxide-inducible genes OxyR [Shigella flexneri 2a str. 301]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGMLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPHIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCVILALVKESEAFIEVPLFDEPMLLAIYEDHPWANRECVLMADLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPPERKRDGVVYLPCIKPEPCRTIGLVYRPGSPLRS
RYEQLAEAIRARMDGHFDKVLKQAV

>gi|150957599|gb|ABR79629.1| activator, hydrogen peroxide-inducible genes [Klebsiella pneumoniae subsp. pneumoniae MGH 78578]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTIGPYLLPHIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCAILALVKESEAFIEVPLFDEPMMLAIYEDHPWANRDRVPMSDLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPQERKRDGVVYLPCIKPEPRRTIGLVYRPGSPLRS
RYEHLAEAIRGTMDGHFDKALKQAV

>gi|16762318|ref|NP_457935.1| hydrogen peroxide-inducible regulon activator [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTIGPYLLPLIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCAILALVKESEAFIEVPLFDEPMMLAIYEDHPWANRDRVPMSDLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPQERKRDGVVYLPCIKPEPRRTVGLVYRPGSPLRS
RYEQLAEAIRGAMDGHFDKALKQAV

>gi|62182584|ref|YP_219001.1| regulatory protein sensor for oxidative stress, regulates intracellular hydrogen peroxide (LysR family) [Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVDQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTIGPYLLPLIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCAILALVKESEAFIEVPLFDEPMMLAIYEDHPWANRDRVPMSDLAGEKLLMLEDGHCLRDQAMGFCFE
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPQERKRDGVVYLPCIKPEPRRTVGLVYRPGSPLRS
RYEQLAEAIRGAMDGHFDKALKQAI

>gi|146313657|ref|YP_001178731.1| transcriptional regulator, LysR family [Enterobacter sp. 638]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVDQART
VLREVKVLKEMASQQGEAMSGPLHIGLIPTVGPYLLPLIIPMLHQTFPKLEMYLHEAQTHQLLAQLDSGK
LDCAILALVKESEAFIEVPLFDEPMMLAIYEDHPWANRDRVPMGDLAGEKLLMLEDGHCLRDQAMGFCFE
AGAEEDTHFRATSLETLRNMVAAGSGITLLPALAVPRERRRDGVVYLPCIKPEPRRTVGLVYRPGSPLRS
RYEQLAEAIRSSMDGHFDDALKQAV

>gi|8134603|sp|P71318|OXYR_PECCC Hydrogen peroxide-inducible genes activator
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVEQART
VLREVKVLKEMASQQGETMSGPLHIGLIPTVGPYLLPQIIPMLHRTFPKLEMYLHEAQTHQLLAQLDSGK
LDCAILAMVKESEAFIEVPLFDEPMKLAIYQDHPWANRERVAMSDLAGEKLLMLEDGHCLRDQAMGFCFQ
AGADEDTHFRATSLETLRNMVAAGSGITLLPALSVPRERERDGVCYLPCYKPEPKRTIALVYRPGSPLRG
RYEQLADTIREHMQGYMETLSK

>gi|50123163|ref|YP_052330.1| hydrogen peroxide-inducible genes activator [Erwinia carotovora subsp. atroseptica SCRI1043]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVEQART
VLREVKVLKEMASQQGESMSGPLHIGLIPTVGPYLLPQIIPMLHRTFPKLEMYLHEAQTHQLLAQLDSGK
LDCAILAMVKESEAFIEVPLFDEPMKLAIYQDHPWANRERVAMSDLAGEKLLMLEDGHCLRDQAMGFCFQ
AGADEDTHFRATSLETLRNMVAAGSGITLLPALSVPRERERDGVCYLPCYKPEPKRTIALVYRPGSPLRA
RYEQLADTIREHMQGYMENLSK

>gi|118069258|ref|ZP_01537505.1| transcriptional regulator, LysR family [Serratia proteamaculans 568]
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVEQART
VLREVKVLKEMASQQGEAMSGPLHIGLIPTVGPYLLPQIIPTLHKTFPKLEMYLHEAQTQQLLAQLDSGK
LDCAILALVKETEAFIEVPLFDEPMKLAVYSDHPWSQRDRVAMPDLAGEKLLMLEDGHCLRDQAMGFCFQ
AGADEDTHFRATSLETLRNMVAAGSGITLLPSLAVPPQRERDGVCYLDCYKPEPKRTIALVYRPGSPLRS
RYEQLAEAIREHMQNYLGTTLKQAV

>gi|85060136|ref|YP_455838.1| hydrogen peroxide-inducible regulon activator [Sodalis glossinidius str. 'morsitans']
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQVRKLEDELGVMLLERTSRKVLFTQAGLLLVDQART
VLREVKVLREMASQQGESMSGPLHIGLIPTVGPYLLSHIVPMLHQAFPKLEMYLHESQTSQLLQQLDSGR
LDCVILALVKESEAFIEVPLFDETMQLAIYADHPWADRDRVPMSDLAGERLLMLEDGHCLRDQAMGFCFQ
AGADEDTHFRATSLETLRNMVAAGSGITLLPSLAVPRERKRDGVCYLPCYRPEPKRTIALVYRPGSPLRA
RYEQLADCIRSHMQAYMGTGLKQAV

>gi|8134604|sp|Q9X725|OXYR_ERWCH Hydrogen peroxide-inducible genes activator
MNIRDLEYLVALAEHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVEQART
VLREVKVLKEMASQQGEAMSGPLHIGLIPTVGPYLLPQIIPMLHRAFPKLEMYLHEAQTHQLLAQLDSGK
LDCAILAMVKESEAFIEVPLFDEPMKLAIYQDHPWANRERVAMSDLSGEKLLMLEDGHCLRDQAMGFCFQ

AGADEDTHFRATSLETLRNMVAAGSGITLLPSLAVPQERIRDGVCYLPCYKPEPKRTIALVYRPGSPLRG
RYEQLADSVREHMQLHMEKLSAQSA

>gi|51594478|ref|YP_068669.1| oxidative stress transcriptional
regulatory protein [Yersinia pseudotuberculosis IP 32953]
MNIRDLEYLVALAEFRHFRRAADSCHVSQPTLSGQIRKLEDELGIMLLERTSRKVLFTQAGLLLVEQAKT
VLREVKVLKEMASLQGESMSGPLHIGLIPTVGPYLLPQIIPMLHKTFPKLEMYLHEAQTQNLLAQLDSGK
LDCAILALVKETEAFIEIPLFDEPMNLAIYADHPWANRERVEMHELAGEKLLMLEDGHCLRDQAMGFCFQ
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPNERQRDGVCYLECYKPVPKRTIALVYRPGSPLRG
RYEQLAEAIRDHMQERMASSLEQAI

>gi|16124047|ref|NP_407360.1| oxidative stress transcriptional
regulatory protein [Yersinia pestis CO92]
MNIRDLEYLVALAEFRHFRRAADSCHVSQPTLSGQIRKLEDELGIMLLERTSRKVLFTQAGLLLVEQAKT
VLREVKVLKEMASLQGESMSGPLHIGLIPTVGPYLLPQIIPMLHKTFPKLEMYLHEAQTQNLLAQLDSGK
LDCAILALVKETEAFIEIPLFDEPMNLAIYADHPWANRERVEMHELAGEKLLMLEDGHCLRDQAMGFCFQ
AGADEDTHFRATSLETLRNMVAAGSGITLLPALAVPNERQRDGVCYLECYKPVPKRTIALVYRPGSPLRG
RYEQLAEAIRDHMQERMAPSLEQAI

>gi|123440532|ref|YP_001004526.1| oxidative stress transcriptional
regulatory protein [Yersinia enterocolitica subsp. enterocolitica 8081]
MNIRDLEYLVALAEFRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVEQART
VLREVKVLKEMASLQGESMSGPLHIGLIPTVGPYLLPQIIPMLHQTFPKLEMYLHEAQTQNLLAQLDSGK
LDCAILALVKETEAFIEVPLFDEPMQLAIYADHPWADRDKVQMHELAGEKLLMLEDGHCLRDQAMGFCFQ
AGADEDTHFRATSLETLRNMVAAGSGITLLPSLAVPNERKRDGVCYLECYKPVPKRTIALVYRPGSPLRG
RYEQLAEAIREHMQPRMNSDVINQKLEQAV

>gi|77961065|ref|ZP_00824913.1| COG0583: Transcriptional regulator
[Yersinia mollaretii ATCC 43969]
MNIRDLEYLVALAEFRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVEQART
VLREVKVLKEMASLQGESMSGPLHIGLIPTVGPYLLPQIIPTLHQAFPKLEMYLHEAQTQNLLAQLDSGK
LDCAILALVKETEAFIEVPLFDEPMQLAIYADHPWADRDRVQMHELAGEKLLMLEDGHCLRDQAMGFCFQ
AGAEEDTHFRATSLETLRNMVAAGSGITLLPSLAVPNERKRDGVCYLECYKPVPKRTIALVYRPGSPLRG
RYEQLAEAIREHMQPRMSEELKQAV

>gi|77973001|ref|ZP_00828555.1| COG0583: Transcriptional regulator
[Yersinia frederiksenii ATCC 33641]
MNIRDLEYLVALAEFRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVEQART
VLREVKVLKEMASLQGESMSGPLHIGLIPTVGPYLLPQIIPMLHQTFPKLEMYLHEAQTQNLLAQLDSGK
LDCAILALVKETEAFIEVPLFDEPMQLAIYSDHPWASREKVQMHELAGEKLLMLEDGHCLRDQAMGFCFQ
AGAEEDTHFRATSLETLRNMVAAGSGITLLPSLAVPNERTRDGVCYLECYKPVPKRTIALVYRPGSPLRG
RYEQLAEAIREHMQPRMSPEVINEKSEQAV

>gi|77976810|ref|ZP_00832280.1| COG0583: Transcriptional regulator
[Yersinia intermedia ATCC 29909]
MNIRDLEYLVALAEFRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVEQART
VLREVKVLKEMASLQGESMSGPLHIGLIPTVGPYLLPQIIPMLHKTFPKLEMYLHEAQTQNLLAQLDSGK
LDCAILALVKETEAFIEVPLFDEPMNLAIYADHPWADRDRVQMHELAGEKLLMLEDGHCLRDQAMGFCFQ
AGAEEDTHFRATSLETLRNMVAAGSGITLLPSLAVPNERKRDGVCYLECYKPVPKRTVALVYRPGSPLRG
RYEQLAEAIREHMQPRMSSEVTNEKLEQAV

>gi|77957411|ref|ZP_00821468.1| COG0583: Transcriptional regulator
[Yersinia bercovieri ATCC 43970]
MNIRDLEYLVALAEFRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQAGLLLVEQART

VLREVKVLKEMASLQGESMSGPLHIGLIPTVGPYLLPQIIPTLHQTFPKLEMYLHEAQTQNLLAQLDSGK
LDCAILALVKESEAFIEVPLFDEPMQLAIYADHPWAGRDRVQMHELAGEKLLMLEDGHCLRDQAMGFCFQ
AGAEEDTHFRATSLETLRNMVAAGSGITLLPSLAVPNERERDGVCYLECYKPIPKRTIALVYRPGSPLRG
RYEQLAEAIREHMQPRMSQDVISEKLEQAV

>gi|37528557|ref|NP_931902.1| hydrogen peroxide-inducible genes
activator [Photorhabdus luminescens subsp. laumondii TTO1]
MNIRDLEYLVALAEHRHFRHAADSCHVSQPTLSGQIRKLEDNLGVMLLERTSRKVLFTQQGMLLVEQART
VLREVRILQEMASLQGENMSGPLHIGLIPTVGPYLLPHIIPRLHSLFPKLEMYLYEAQTQSLLAQLDSGK
LDCAILAMVKETRDFIEVPLFEEPMKLAIYDDHPWAERKKIAMDELAGEKLLMLEDGHCLRDQAMGFCFQ
AGAKEDTHFRATSLETLRNMVAAGSGITLLPDLAVPQEKKRDGVCYLECNSPEPKRSVVLIYRPGSPLRG
RYEQLAEAIRAKMGTYYGQSK

>gi|53733092|ref|ZP_00349646.1| COG0583: Transcriptional regulator
[Haemophilus influenzae R2846]
MNIRDLEYLVALSEYKHFRRAADSCNVSQPTLSGQIRKLEDELGIILLERTSRKVLFTQSGMLLVDQART
VLREVKLLKEMASNQGKEMTGPLHIGLIPTVGPYLLPYIVPMLKAAFPDLEVFLYEAQTHQLLEQLETGR
LDCAIVATVPETEAFIEVPIFNEKMLLAVSEHHPWAQESKLPMNQLNGQEMLMLDDGHCLRNQALDYCFT
AGAKENSHFQATSLETLRNMVAANAGITFMPELAVLNEGTRKGVKYIPCYSPEPSRTIALVYRPGSPLRN
RYERVASAVSDEVKSILGGLK

>gi|16272514|ref|NP_438728.1| hydrogen peroxide-inducible genes
activator [Haemophilus influenzae Rd KW20]
MNIRDLEYLVALSEYKHFRRAADSCNVSQPTLSGQIRKLEDELGIILLERTSRKVLFTQSGMLLVDQART
VLREVKLLKEMASNQGKEMTGPLHIGLIPTVGPYLLPYIVPMLKAAFPDLEVFLYEAQTHQLLEQLETGR
LDCAIVATVPETEAFIEVPIFNEKMLLAVSEHHPWAQESKLPMNQLNGQEMLMLDDGHCLRNQALDYCFT
AGAKENSHFQATSLETLRNMVAANAGITFMPELAVLNEGTRKGVKYIPCYSPEPSRTIALVYRPGSPLRN
RYERVASAVSDEVKSILDGLK

>gi|145631749|ref|ZP_01787510.1| DNA-binding transcriptional regulator
OxyR [Haemophilus influenzae 22.4-21]
MNIRDLEYLVALSEYKHFRRAADSCNVSQPTLSGQIRKLEDELGIILLERTSRKVLFTQSGMLLVDQART
VLREVKLLKEMASNQGKEMTGPLHIGLIPTVGPYLLPYIVPMLKAAFPDLEVFLYEAQTHQLLEQLETGR
LDCAIVATVPETEAFIEVPIFNEKMLLAVSEHHPWAQESKLPMNQLNGQEMLMLDDGHCLRNQTLDYCFT
AGAKENSHFQATSLETLRNMVAANAGITFMPELAVLNEGTRKGVKYIPCYSPEPSRTIALVYRPGSPLRN
RYERVASAVSDEVKSILGGLK

>gi|145637931|ref|ZP_01793573.1| hydrogen peroxide-inducible genes
activator [Haemophilus influenzae PittHH]
MNIRDLEYLVALSEYKHFRRAADSCNVSQPTLSGQIRKLEDELGIILLERTSRKVLFTQSGMLLVDQART
VLREVKLLKEMASNQGKEMTGPLHIGLIPTVGPYLLPYIVPMLKAAFPDLEVFLYEAQTHQLLEQLETGR
LDCAIVATVPETEAFIEVPIFNEKMLLAVSEHHPWAQESKLPMNQLNGQEMLMLDDGHCLRNQALDYCFT
AGAKENSHFQATSLETLRNMVAANAGITFMPELAVLNEGTRRGVKYIPCYSPEPSRTIALVYRPGSPLRN
RYERVASAVSDEVKSILDGLK

>gi|145635342|ref|ZP_01791044.1| DNA-binding transcriptional regulator
OxyR [Haemophilus influenzae PittAA]
MNIRDLEYLVALSEYKHFRRAADSCNVSQPTLSGQIRKLEDELGIILLERTSRKVLFTQSGMLLVDQART
VLREVKLLKEMASNQGKEMTGPLHIGLIPTIGPYLLPYIVPTLKAAFPDLEVFLYEAQTHQLLEQLETGR
LDCAIVATVPETEAFIEVPIFNEKMLLAVSEHHPWAQESKLPMNQLDGQEMLMLDDGHCLRNQALDYCFT
AGAKENSHFQATSLETLRNMVAANAGITFMPELAVLNEGTRKGVKYIPCYSPEPSRTIALVYRPGSPLRN
RYERVASAVSDEVKSILGGLK

>gi|145640547|ref|ZP_01796131.1| hydrogen peroxide-inducible genes activator [Haemophilus influenzae R3021]
MNIRDLEYLVALSEYKHFRRAADSCNVSQPTLSGQIRKLEDELGIILLERTSRKVLFTQSGMLLVDQART
VLREVKLLKEMASNQGKEMTGPLHIGLIPTVGPYLLPYIVPMLKAAFPDLEVFLYEAQTHQLLEQLETGR
LDCAIVATVPETEAFIEVPIFNEKMLLAVSEHHPWAQESKLLMNQLNGQEMLMLDDGHCLRNQTLDYCFT
AGAKENSHFQATSLETLRNMVAANAGITFMPELAVLNEGTRKGVKYIPCYSPEPSRTIALVYRPGSPLRN
RYERVASAVSDEVKSILDGLK

>gi|38488592|dbj|BAD02310.1| oxyR like protein [Actinobacillus actinomycetemcomitans]
MNIRDLEYLVALAEHKHFRRAADSCHVSQPTLSGQIRKLEDELGIILLERTSRKVLFTQSGLLLVDQAKT
VLREVKLLKEMASNQGKDMTGPLHIGIIPTVGPYLMPHIVPTLQQNFPDLELFLYEAQTYRLLEQLETGR
LDCAIVASVPETEAFIEVQLFNEKMLLAVAEQHPWANENSVSMSLLKDCEILMLDDGHCLRNQALGYCFT
AGARENAHFQATSLETLRNMVAANTGVTLMPQLAVLSEGNRSGVKYLPCDEPEPSRDITLVYRPGSPLRA
RYERVANTVSQSVKSILSS

>gi|15603211|ref|NP_246285.1| OxyR [Pasteurella multocida subsp. multocida str. Pm70]
MNIRDLEYLVALAEHKHFRRAADACHVSQPTLSGQIRKLEDELGIILLERTSRKVLFTQSGLLLVDQAKN
VLKEVKLLKEMASNQGKDMTGPIHIGVIPTVGPYLLPYIMPVLKETFPDLELFLYEAQTNQLLEQLETGH
LDCAIVASVRETEAFIEVPIFHEAMLLAVSENHPWANERTIAMNRLNGCEMLMLDDGHCLRDQTIGYCFS
AGAKENAHFQATSLETLRNMVASNTGITLMPKLAVINEGNRTGVKYIPCHSPAPSRAITLVYRPGSPLRN
RYEKIAQTISHSVQDVLD

>gi|90413601|ref|ZP_01221591.1| putative Hydrogen peroxide-inducible genes activator [Photobacterium profundum 3TCK]
MNIRDLEYLIALSEHKHFRKAAESCFVSQPTLSGQIRKLEDELGVSLLERTSRRVLFTDAGLSLVAQAQK
VLLEVKVLTELASQQGESMSGPLHIGFIPTVGPYLLPLIIPMFRESFPDLELFLHEAQTSQLTHLLEEGK
LDCILLAAVKETESFIELPLYDEPMVVAVPDTHPWAEKDEMDMASLHGETLLMLGDGHCLRDQAMGFCFA
AGANEDGRFKATSLETLRNMVAAGSGITLLPQLATPKERSRDGVCYIKVHDPIPTRLITLCYRPGSPLRT
RYEKIATEIKDRMVTYFEQ

>gi|75437399|ref|ZP_00733420.1| transcriptional regulator [Actinobacillus succinogenes 130Z]
MNIRDLEYLVSLAEFKHFRRAADACNVSQPTLSGQIRKLEDELGITLLERTSRKVLFTQSGLLLVAQAKQ
VLREVKLLKEMASNQGKDMTGPLHVGVIPTVGPYLLPYIMPVLKESFPELELFLYEAQTNQLVDQLETGR
LDCAIVAMVVETEPFIQVPLFNEKMLIAVSEAHPWAKEKNIPLDYLKGTEVLMLDDGHCLREQALGYCFA
AGASENSHFQATSLEMLRNMIAANAGVTLMPELAVLNEGQRRGVKYIPCINPEPQRTIALIYRPGSPLRA
RYERVANAVKKAVRPILEGD

>gi|52424209|ref|YP_087346.1| LysR protein [Mannheimia succiniciproducens MBEL55E]
MNIRDLEYLAALAEYKHFRRAADACHVSQPTLSGQIRKLEDELGITLLERTSRKVLFTQSGLILVEQAKK
VLREVKLLKEMASNQGKEMTGPLHLGVIPTVGPYLLPYIMPALKEAFPDLELYLYEAQTSHLLDQLESGR
LDCAILATVPETEPFIEVPIFNERMLLAVSEQHPWAKEKSIKMHALQGHEVLMLDDGHCLRDQALGYCFT
AGARENSHFQATSLETLRNMIAANAGMTLMPELAMLNEGTRAGVKYIPCTDPEPKRTIALVYRPGSPLRS
RYERVANAVGDAVKAILHTEGD

>gi|54307494|ref|YP_128514.1| putative Hydrogen peroxide-inducible genes activator [Photobacterium profundum SS9]
MNIRDLEYLIALSEHKHFRKAAESCFVSQPTLSGQIKKLENELGVSLLERTSRRVLFTDAGLTLVAQAQK
VLLEVKVLTELASQQGESMSGPLHIGFIPTVGPYLLPLIIPMFRESFPDLELFLHEAQTSQLTHLLEEGK
LDCILLAAVKETESFIELPLYDEPMVVAVPDTHPWAEKDEMDMASLHGETLLMLGDGHCLRDQAMGFCFA
AGAKEDGRFKATSLETLRNMVAAGSGITLLPQLATPKERSRDGVCYIKVHDPIPTRLITLCYRPGSPLRT

RYEKIAAEIKDRMVTYFEQ

>gi|90580866|ref|ZP_01236668.1| putative Hydrogen peroxide-inducible genes activator [Vibrio angustum S14]
MNIRDLEYLVALSEHKHFRKAAEACYVSQPTLSGQIRKLEDELGVSLLERTSRRVLFTDAGLSLVAQAQK
VLLEVKILTELASVQGESMSGPLHIGFIPTVGPYLLPQIIPSLKEAFPELELFLHEAQTHQLVQQLEEGK
LDCIILAAVKESEPFIELPLYDEPMMLAVPETHKWASEKDIDMSLLHGESLLMLEDGHCLRNQALGFCFA
AGARDDGRFKATSLETLRNMVAAGSGITLLPQLASPKEHCRDGVCYIAAKHPQPTRLITLAYRPGSPLKA
RYEKLAEVIKEKMPEVFAKHTQP

>gi|46156415|ref|ZP_00132876.2| COG0583: Transcriptional regulator [Haemophilus somnus 2336]
MNIRDLEYLVSLAEHKHFRRAADACYVSQPTLSGQIRKLEDELGIVLLERTSRKVLFTQSGLLLVKQAKT
VLREIKLLKEMASNQGKEMNGPLHMGVIPTIGPYLLPYIVPALKNTFPDLELFLYEAQTQKLLEQLETGH
LDCVILASVDEAEAFIEVPMFNERMLLAVSDEHPLSKEDSIKMDKLKGYEMLMLDDGHCLRNQALDYCFA
AGAKENQNFRATSLETLRNMVSANTGITLIPELALLNEGSRKGIKYLPCFSPEPSRGISLVYRPGSPLRG
RYERIANKVSEIIKPLLNNRKNGN

>gi|89075201|ref|ZP_01161632.1| putative Hydrogen peroxide-inducible genes activator [Photobacterium sp. SKA34]
MNIRDLEYLVALSEHKHFRKAAEACYVSQPTLSGQIRKLEDELGVSLLERTSRRVLFTDAGLSLVAQAQK
VLLEVKILTELASVQGESMSGPLHIGFIPTVGPYLLPQIIPSLKEAFPELELFLHEAQTHQLVQQLEEGK
LDCMILAAVKESEPFIELPLYDEPMMLAVPETHKWASEKDIDMSLLQGESLLMLEDGHCLRNQALGFCFA
AGARDDGRFKATSLETLRNMVAAGSGITLLPQLASPNEHCRDGVCYIAAKHPQPTRLITLAYRPGSPLKA
RYEKLAEVIKEKMPEVFAKHTQP

>gi|59712906|ref|YP_205682.1| hydrogen peroxide-inducible genes activator [Vibrio fischeri ES114]
MNIRDFEYLVALAEHKHFRKAAESCFVSQPTLSGQIKKLEDEVGLALLERTSRKVLFTEAGLQLVEQAKK
ILLEVKRFSELANQQGKEMTGPLHLGFIPTVGPYVLPWIVPTLKAQFPDLNLYLHEAQTHQLVKMLEEGK
IDCMILASVEETNMFIEVPVYDEPMVLAVPKDHKWAKEVSIDMSRLSGESVLMLGDGHCLRDQALGFCFA
AGAKDDDHFKATSLETLRNMVAAGGGITLMPYLSVPKEKERDGVCYLPAQDPVPHRQIVLAYRPGSPLRA
RYESLAKEIENKMSNVIRS

>gi|86147322|ref|ZP_01065636.1| transcriptional regulator, LysR family protein [Vibrio sp. MED222]
MNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEDEIGLQLTERSPRKVIFTESGLQLVEQAKR
ILNEVKTFKDMASGHGEAMTGPMHIGFIPTVGPYILPKIIPHLKDSFPDLELYLHEAQTHQLVSQLEDGK
LDCLVLAAVDETAAFKEIDVYNEPLSVAVPCDHEWAKQDTVDMLQLNGQTVLALGDGHCLRDQALGFCFA
AGAKDDERFKATSLETLRNMVAAGAGITLLPQLSVPKEKQKDGVCYVPAVNPTPSRRIVVAYRPGSPLKG
RFEQLAEAIRTQLDKAV

>gi|84390825|ref|ZP_00991517.1| transcriptional regulator, LysR family [Vibrio splendidus 12B01]
MNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEDEIGLQLTERSPRKVIFTESGLQLVEQAKR
ILNEVKTFKDMASGHGEAMTGPMHIGFIPTVGPYILPKIIPHLKENFPDLELYLHEAQTHQLVSQLEDGK
LDCLVLAAVDETAAFKEIDVYDEPLSVAVPCDHEWAQQDAVDMLQLNGQTVLALGDGHCLRDQALGFCFA
AGAKDDERFKATSLETLRNMVAAGAGITLLPQLSVPKEKQKDGVCYVPAVNPTPSRRIVVAYRPGSPLKG
RFEQLAEAIRTQLEKTA

>gi|148979177|ref|ZP_01815356.1| DNA-binding transcriptional regulator OxyR [Vibrionales bacterium SWAT-3]
MNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEDEIGLQLTERSPRKVIFTESGLQLVEQAKR
ILNEVKTFKDMASGHGEAMTGPMHIGFIPTVGPYILPKIVPHLKESFPELELYLHEAQTHQLVSQLEDGK

LDCLVLAAVDETAVFKEIDVYDEPLSVAVPCDHEWAQQDTVDMLQLNGQTVLALGDGHCLRDQALGFCFA
AGAKDDERFKATSLETLRNMVAAGAGITLLPQLSVPKEKQKDGVCYVPAVNPTPSRRIVVAYRPGSPLKG
RFEQLAETIRTQLEKVV

>gi|149189164|ref|ZP_01867451.1| DNA-binding transcriptional regulator OxyR [Vibrio shilonii AK1]
MNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEDEIGTSLLERSSRRVLFTDAGLQLVEQAKR
VLKEVKTFREMAAGQSGEMTGPMHIGFIPTVGPYLLPKIIPKLKDAFPELELYLHEAQTHQLVRQLEEGK
LDCLVLASVPETAPFKEIEVYNEPMSVAVPCDHEWANKDQIEMAELNGQTVLSLGDGHCLRDQALGFCFA
AGARDDERFKATSLETLRNMVAAGAGITLLPELSLPTEKVKDGVCYVTAINPTPSRSIVLAYRPGSPLRA
RFEKLAKAITEYLS

>gi|148868418|gb|EDL67529.1| hydrogen peroxide-inducible genes activator [Vibrio harveyi HY01]
MNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEDELGTALLERSSRRVLFTDSGLQLVEQAKR
ILSEVKTFKDMASGQSGAMTGPMHIGFIPTVGPYLLPKIVPRLKEEFPELELFLHEAQTHQLVRQLEEGK
LDCLVLASVDETAPFKEIEVYNEPLSVAVPCDHEWAGLDHIDMLDLNGRTVLALGDGHCLRDQALGFCFA
AGAKDDERFKATSLETLRNMVAAGAGITLLPELSVPHEKKKDGVCYVPAVNPTPSRSIVLVYRPGSPLRA
RFEALASTIKSILEAKQNSIAA

>gi|116183621|ref|ZP_01473601.1| hypothetical protein VEx2w_02003825 [Vibrio sp. Ex25]
MNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEDELGTTLLERSSRRVLFTDSGLQLVEQAKR
ILSEVKTFKDMASGQSGAMTGPMHIGFIPTVGPYLLPKILPQIKEAFPELELFLHEAQTHQLVRQLEEGK
LDCLVLASVDETAPFKEIEVYNEPLSVAVPCDHEWASLDHVDMLELNGKTVLALGDGHCLRDQALGFCFA
AGAKDDERFKATSLETLRNMVAAGAGITLLPQLSIPNEKQKDGVCYVPAVNPTPSRNIVLVYRPGSPLRA
RFEALAAKIKEVLASYPSLNAAA

>gi|150424013|gb|EDN15952.1| transcriptional regulator, LysR family [Vibrio cholerae AM-19226]
MNIRDFEYLVALADHKHFRKAAEACFVSQPTLSGQIRKLEDEIGTTLLERSSRRVLFTEAGLQLVDQAKK
ILSEVKTFKDMANQQTGAMTGPLHIGFIPTLGPYLLPKIIPTLKERFPELELYLHEAQTNQLVRQLEEGK
LDCLVLASVEETAPFKEIELYNEVLSIAVPCDHAWAARDEVDMLELKGKTVLALGDGHCLRDQALGFCFA
AGAKDDERFKATSLETLRNMVAAGAGITLLPELALPEDKTKDGVCYLRAINPIPSRRLVLAYRPGSPLRQ
RFEQLAEVIKHRLQQSE

>gi|15642631|ref|NP_232264.1| transcriptional regulator, LysR family [Vibrio cholerae O1 biovar eltor str. N16961]
MNIRDFEYLVALADHKHFRKAAEACFVSQPTLSGQIRKLEDEIGTTLLERSSRRVLFTEAGLQLVDQAKK
ILSEVKTFKDMANQQTGAMTGPLHIGFIPTLGPYLLPKIIPTLKERFPELELYLHEAQTNQLVRQLEEGK
LDCLVLASVEETAPFKEIELYNEVLSIAVPCDHAWAARDEVDMLELKGKTVLALGDGHCLRDQALGFCFA
AGAKDDERFKATSLETLRNMVAAGAGITLLPELALPEDKTKDGVCYLRAVNPIPSRRLVLAYRPGSPLRQ
RFEQLAEVIKHRLQQSE

>gi|28899526|ref|NP_799131.1| transcriptional regulator, LysR family [Vibrio parahaemolyticus RIMD 2210633]
MNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEDELGTALLERSSRRVLFTDSGLQLVDQAKR
ILSEVKTFKDMASGQSGAMTGPMHIGFIPTVGPYLLPKILPQLKEEFPELELFLHEAQTHQLVRQLEEGK
LDCLVLASVAETAPFKEIEVYNEPLSVAVPCGHEWAQLDQVDMLELNGKTVLALGDGHCLRDQALGFCFA
AGAKDDERFKATSLETLRNMVAAGAGITLLPQLSIPAEKQKDGVCYIPAVNPTPSRSIVLAYRPGSPLRA
RFEALAAKIKAILESQPSSMAA

>gi|91227290|ref|ZP_01261715.1| transcriptional regulator, LysR family protein [Vibrio alginolyticus 12G01]

MNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEDELGTTLLERSSRRVLFTDSGLQLVEQAKR
ILSEVKTFRDMASGQSGAMTGPMHIGFIPTVGPYLLPKILPQIKEAFPELELFLHEAQTHQLVRQLEEGK
LDCLVLASVDETAPFKEIEVYQEPLSVAVPCDHEWASLDHVDMLELNGKTVLALGDGHCLRDQALGFCFA
AGAKDDERFKATSLETLRNMVAAGAGITLLPQLSIPNEKQKDGVCYVPAVNPTPSRNIVLVYRPGSPLRA
RFEALAAKIKEVLASYPSLNAAA

>gi|46143704|ref|ZP_00204557.1| COG0583: Transcriptional regulator
[Actinobacillus pleuropneumoniae serovar 1 str. 4074]
MNIRDLEYLIALADYKHFRRAADACNVSQPTLSGQIRKLEDELGTVLLERTSRKVLFTQAGLTLVEQAKA
VLREVKVLKEMASNQGKEMSGPLHVGIIPTLGPYLLPLVLPALKSTFPELELYIYELQTTQLVDQLESGQ
LDCGILAFVKESEPFIEVPIFNEQMLLAVSDKHEWSHKSKMDISYLKDKELLFLDDGHCLRTQTLDYCLS
VGAKESTHFKATNLETLRNMVAANVGMSLIPELAAKPCEGLNYLTFDEPKPYRTVGLIYRPGSPLRIRYE
RLAKEVSKIMKQEKIHE

>gi|27364769|ref|NP_760297.1| Transcriptional regulator [Vibrio
vulnificus CMCP6]
MNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEEEIGTTLLERSSRRVLFTDAGLQLVEQAKN
ILKEVKTFKEMASGQSGAMTGPMHIGFIPTLGPYLLPKIVPQLKERFPELELFLHEAQTQQLVRQLEDGK
LDCLVLASVAETEPFKEIEVYNEPLSVAVPCHHEWAALEQLDMLELNGKTVLALGDGHCLRDQALGFCFA
AGARDDERFKATSLETLRNMVAAGAGITLLPELSVPKEKQKDGVCYIKAVNPVPSRTIVVVYRPGSPLRA
RFEQLAATIKELLVSGSEQ

>gi|37681180|ref|NP_935789.1| transcriptional regulator [Vibrio
vulnificus YJ016]
MGMNIRDFEYLVALAEHKHFRKAAEACFVSQPTLSGQIRKLEEEIGTTLLERSSRRVLFTDAGLQLVEQA
KNILKEVKTFKEMASGQSGAMTGPMHIGFIPTLGPYLLPKIVPQLKERFPELELFLHEAQTQQLVRQLED
GKLDCLVLASVAETEPFKEIEVYNEPLSVAVPCHHEWAALEQLDMLELNGKTVLALGDGHCLRDQALGFC
FAAGARDDERFKATSLETLRNMVAAGAGITLLPELSVPKEKQKDGVCYIKAVNPVPSRTIVVVYRPGSPL
RARFEQLAVTIKELLVSGSEQ

>gi|33151874|ref|NP_873227.1| hydrogen peroxide-inducible genes
activator [Haemophilus ducreyi 35000HP]
MNIRDLEYLIALADYKHFRRAADACNVSQPTLSGQIRKLENELGTILLERTSRKVLFTQAGLTLVEQAKA
VLREVKILKEMASNQGKEMSGPLHVGIIPTLGPYLLPFALPALKSAFPELDLYIYELQTSQLIDQLEAGQ
LDCGILALVKESEPFIEIPIFNEEMLLAVPKQHEWAKQSSLTINALKDKELLFLDDGHCLRTQTLDYCLS
VGAKESTHFKATNLETLRNMVATNAGMSLVPELAAKQNANIHYLTFENPQPYRAIGLIYRPGSPLRIRYE
RLAKEVAYIMTKEGKNE

>gi|117620885|ref|YP_857687.1| hydrogen peroxide-inducible genes
activator [Aeromonas hydrophila subsp. hydrophila ATCC 7966]
MRARILDPGPPCFIGLSTFPLAPSLVRSAVLGHANGPIGWQDSGCMNLRDLEYLVALEEEKHFRKAAERC
FVSQPTLSGQLRKLEDELGVILIERTSRKVLFTPAGDAMAQQARKVLKEVRELKSIGQHFAEPMSGEIHI
GFIPTVGPYLLPHIIQDLREHFPKLEFYLYEEQTQVLLKRLEEGELDCLILAELEGMDGFGSIPLYQEPM
WLAVPQQHPEAKAKAVPLSNLKGKKLLMLADGHCLRDQAMGFCFAAGIGEDQRFKGTSLETLRNMVAAGS
GMTLVPRLAVPANAEEGGVSYRPVIDPVPGRTIALLYRHYSVRRPCFNELAARISRLMQSLLG

>gi|145298146|ref|YP_001140987.1| hydrogen peroxide-inducible genes
activator [Aeromonas salmonicida subsp. salmonicida A449]
MSLRDLEYLVALEEEKHFRKAAERCFVSQPTLSGQLRKLEDELGVILIERTSRKVLFTPAGDAMAHQARK
VLKEVRELKNIGQHFAEPMSGEIHIGFIPTVGPYLLPHIIQDLREHFPKLEFYLYEEQTQLLLKRLEEGE
LDCLILAELDGMEGFGSIPLYQEPMWLAVPQHHPEAKARAVPLSNLKGKKLLMLADGHCLRDQAMGFCFA
AGIGEDQRFKGTSLETLRNMVAAGSGMTLVPRLAVPANAEEGGVSYRPVVDPVPGRTISLLYRHYSVRRP
CFNELASRISTLMKSLLG

>gi|149909427|ref|ZP_01898082.1| regulatory protein sensor for oxidative stress, regulatesintracellular hydrogen peroxide (LysR family) [Moritella sp. PE36]
MNLRDLEYLVALQELKHFRKAAEKCFVSQPTLSGQIRKLEDELDVILIERTSRKVLFTPAGDQIADQART
VLLESKAIKEIAKSYASPTAGAIHIGLIPTVAPYLLPLIVPSMKKKFPDLDMFLHENQTHELLKQLDEGE
LDCLLLAYLPGMEKYGHIELYKEPLELIIPSSHRFKGRDRVDLSELRGEKVLMLEDGHCLRDQAMDYCFT
AGAEEDQSFKATSLETLRHMIAAEAGVTLLPHLAIPRSRFTEGVEYIKFVEPEPIRKIVLLYRKGSVRRP
CFNDIAEVISKQVAATIV

>gi|90407841|ref|ZP_01216017.1| putative Hydrogen peroxide-inducible genes activator [Psychromonas sp. CNPT3]
MNFRDLEYLIALEELKHFRKAAEKCFVSQPTLSGQIRKLEEELNVQLMERSSRKVIFTQAGLDIVAKAKN
ILVEAKSLREIAKSHNQPMHGQLHIGLIPTVAPYLLPLIIPSIRKEFPDLEVFLHENQTKVLLKQLESGE
LDCLMLALLPDMQAFHNYPLYVEPLELALSETHQWANEHQIDIKKLSGERVLMLADGHCLRDQALGFCFA
AGAIEDNSFKATSLETLRHMIGADNGLTLLPQLAIPLNRHQAGIKYIPFMAPIPTRSIVLLCRKNSVRTQ
CFEQLSTLITSKVNKQLKMY

>gi|119947099|ref|YP_944779.1| transcriptional regulator, substrate-binding, LysR family protein [Psychromonas ingrahamii 37]
MNFRDLEYLIALEELKHFRKAAEKCFVSQPTLSGQIRKLEDELGIQLMERSPRKVLFTPAGLDIVAKAKT
ILLEVKSLKEIAKSYNEPMQGTLHIGLIPTVAPYLLPLIVAVIKANFPDLSLYLYEKQTNLLLKQLEEGE
LDCLILALLPGMESFTQYHLYQEPLELAITDVHPWAKQPEIELNGLRGEHVLMLEDGHCLRDQTKGFCFA
AGALEDGSFQATSLETLRHMISAENGMTLLPQLAIPVNRHEGGIQYIPFKNPKPTREISLLCRKNSVRKI
CFEQLAKLISTTVQAKLKEYG

>gi|116216130|ref|ZP_01482018.1| hypothetical protein VchoR_02002081 [Vibrio cholerae RC385]
MNIRDFEYLVALADHKHFRKAAEACFVSQPTLSGQIRKLEDEIGTTLLERSSRRVLFTEAGLQLVDQAKK
ILSEVKTFKDMANQQTGAMTGPLHIGFIPTLGPYLLPKIIPTLKERFPELELYLHEAQTNQLVRQLEEGK
LDCLVLASVEETAPFKEIELYNEVLSIAVPCDHAWAARDEVDMLELKGKTVLALGDGHCLRDQALGFC

>gi|71278569|ref|YP_271362.1| hydrogen peroxide-inducible genes activator [Colwellia psychrerythraea 34H]
MIKLRDLEYLTAIDKHKHFGKAAQSCFVSQPTLSGQLMKLEEQLGLQLVERHRRNVMLTPAGEQLVKEAR
KVLQAAGQFESCAKALLDPFAGDLHLGLIPTLAPYLLPHIMADLNKALPNINFFLHENQTKVLLQELDEG
KLDVLILPYLDEMDKFESYQLFDEPLMLATPKNHRLANKKDLSLSDLHDEKILTLADGHCLKDQAMGYCF
SAGAKEDNSFQATSLETLRHMVASGMGITLLPALAAQGNLASDTIHYGQFQAPVPVRGISLVIRPNYSRM
QCVRSIVASVRKSLNGIIT

>gi|21230306|ref|NP_636223.1| oxidative stress transcriptional regulator [Xanthomonas campestris pv. campestris str. ATCC 33913]
MNLRDLKYLVALADHKHFGRAATACFVSQPTLSTQIKKLEDELGVPLVERAPRKVMLTPAGREAAMRARS
IVAEVEQMKEAARRSQDPEAGTVRLGIFPTLAPYLLPHVVPRIRERFPRLELLLIEEKSDQLIHQLREGR
MDAALLALPLQDEQLHAEFLFEEPFVLAVPEGHPLSRHDSMTLDDLSEQRLLLLEDGHCLREQALDVCHL
AGALEKSEFQATSLETLRQMVAANVGVTLLPMLAVKPPVARSENIRLIRFREDKQPNRRIAMAWRRSSAM
TAFLEQLSQIFKELPDSLFTLDQPASGPKAVAA

>gi|119877904|ref|ZP_01644878.1| transcriptional regulator, LysR family [Stenotrophomonas maltophilia R551-3]
MNLRDLKYLVALADHKHFGRAAASCFVSQPTLSTQIRKLEEELGLPLVERAPRKVMLTPAGQEAAARARV
IVSEVEQLKEAARRSRDPEAGTVRLGIFPTLGPYLLPHVIPRIRERFPELELLLVEEKSDVLLDRLREGK
LDAALLALPVIDDQLHAEFLFEEPFLLAVSGRHPLARREHLDVQELATQKLLLLEDGHCLRDQALEVCRL
FGANEKSEFRATSLETLRQMVAADVGITLLPSLSVQPPVPRSSNIRLLDFTGEGRPSRRIAMIWRRSSAM
NDFLTELADQFKRLPEALFTLEAVNAGGDASTLPGPVLNG

>gi|2098748|gb|AAC45427.1| oxidative stress transcriptional regulator; OxyR [Xanthomonas campestris]
MNLRDLKYLVALADHKHFGRAASACFVSQPTLSTQIKKLEDELGVSLVERAPRKVMLTPAGREAAVRARS
IVAEVEQMKEAARRSQDPEAGTVRLGIFPTLAPYLLPHVVPRIRQRFPRLELLLIEEKSDQLMHQLREGR
LDAALLALPLQDDQLHAEFLFEEPFVLAVPEGHPLSRHDSMTLDDLSEQRLLLLEDGHCLRDQALDVCHL
AGALEKSEFQATSLETLRQMVAANVGVTLLPLLAVKPPVARSENIRLIRFREDKQPSRRIAMAWRRSSAM
TAFLEQLSQLFKELPESLFTLDQPATGPKAVAA

>gi|78046497|ref|YP_362672.1| transcriptional regulator, LysR family [Xanthomonas campestris pv. vesicatoria str. 85-10]
MNLRDLKYLVALADHKHFGRAASACFVSQPTLSTQIKKLEDELGVSLVERAPRKVMLTPAGREAAVRARS
IVAEVEQMKEAARRSQDPEAGTVRLGIFPTLAPYLLPHVVPRIRERFPRLELLLIEEKSDQLMHQLREGR
LDAALLALPLQDDQLHAEFLFEEPFVLAVPEGHPLSRHDSMTLDDLSEQRLLLLEDGHCLRDQALDVCHL
AGAMEKSEFQATSLETLRQMVAANVGVTLLPLLAVKPPVARSENIRLIRFREDKQPSRRIAMAWRRSSAM
TAFLEQLSQLFKELPESLFTLDQPATGPKAVAA

>gi|21241675|ref|NP_641257.1| oxidative stress transcriptional regulator [Xanthomonas axonopodis pv. citri str. 306]
MNLRDLKYLVALADHKHFGRAASACFVSQPTLSTQIKKLEDELGVSLVERAPRKVMLTPAGREAAVRARS
IVAEVEQMKEAARRSQDPEAGTVRLGIFPTLAPYLLPHVVPRIRERFPRLELLLIEEKSDQLMHQLREGR
LDAALLALPLQDDQLHAEFLFEEPFVLAVPESHPLSRHDSMTLDDLSEQRLLLLEDGHCLRDQALDVCHL
AGALEKSEFQATSLETLRQMVAANVGVTLLPLLAVKPPVARSENIRLIRFREDKQPSRRIAMAWRRSSAM
TAFLEQLSQLFKELPESLFTLDQPAPGPKAVAA

>gi|54294702|ref|YP_127117.1| Hydrogen peroxide-inducible genes activator [Legionella pneumophila str. Lens]
MNLRDLHYFVILADVKHFGEAAKRCFVSQPTLSMQIKKLEEELGVVLFERTNKQVLLTDQGSKLLDRTRK
ILILIDEMKELARQSEDPFTGELRLGVIPTVSPYMLPLVMPELKNEYPRLKVWLIEDQTHRLITKLEQGE
LDVAIMALPIDKRFSCQILYEEKFYFACANTHPLAQAKSVNINDLKNQPIMLLEEGHCLREQAMAVCQSA
KADDIADFTATSLETLRLMVQAGMGVTLLPALSTLTASTNHLKCIPFSEPAPSRIVGLFWRAGTPRQICF
NAIAELITKNVQSKLA

>gi|119469939|ref|ZP_01612744.1| hypothetical protein ATW7_04854 [Alteromonadales bacterium TW-7]
MNLKDFEYVKAIAQHKHFRKAADACFVSQPTLSGQVKKLEQTLGVTLFDRSTKQVTLTAKGVRLLAQIEV
ILEQTQILKELASASNEPLQGKITIGIIPTIAPYLLPTLLTSMKEAFVDSQFAFIEMQTATILDALNNGE
LDFAILADVAELNHYHTIPLYKEDFLVAVSKDNALAKHKKVALSDLQGCSLLMLSDGHCFKDQAQKFCFS
AGVDVSNQYKGNSLETLLALVAMDDGVTFVPKLACTQRTGIDYMPIFPNQQRNVVFACRKHYPHLAGVEQ
LGEWLSAHPNLKTQLAKAL

>gi|52842042|ref|YP_095841.1| hydrogen peroxide-inducible genes activator OxyR [Legionella pneumophila subsp. pneumophila str. Philadelphia 1]
MNLRDLHYFVILADVKHFGEAAKRCFVSQPTLSMQIKKLEEELGVVLFERTNKQVLLTDQGSKLLDRTRK
ILILIDEMKELARQSEDLFTGELRLGVIPTVSPYMLPLVMPELKNEYPRLKVWLIEDQTHRLITKLEQGE
LDVAIMALPIDKRFSCQTLYEEKFYFACANTHPLAQAKSVNINDLKNQPIMLLEEGHCLREQAMAVCQSA
KADDIADFTATSLETLRLMVQAGMGVTLLPALSTLTASTNHLKCIPFSEPAPSRILGLFWRAGTPRQVCF
NAIAELITKNVQSKLA

>gi|15838133|ref|NP_298821.1| oxidative stress transcriptional regulator [Xylella fastidiosa 9a5c]
MWNYIPSLAARFGFMNLRDLKYLIALADYKHFGRAATACFVSQPTLSTQIKKLEGELGVSLVERAPRKVM
MTPAGREAAIRARSIVAEVEEMKEAARRSRDPEAGAVRLGIFPTLGPYLLPHVVPSIRYRFPQLELLLVE

EKSDELLAQLREGKLDAALLALPLHDEQLHTEFLFEEPFVLAVPEGHPLATRREMTMEELADERLLLLQD
GHCLREQALDVCHMTGASEKSEFQATSLETLRQMVVANVGITLLPLLSVKPPVVCSESIRLINFPLDKQP
SRRIAMVWRRSSAMTTFLERFSGMFKELPKELFDLPQTVVLYKGR

>gi|71274705|ref|ZP_00650993.1| regulatory protein, LysR:LysR, substrate-binding [Xylella fastidiosa Dixon]
MNLRDLKYLIALADYKHFGRAATACFVSQPTLSTQIKKLEGELGVSLVERAPRKVMMTPAGREAAIRARS
IVAEVEEMKEAARRSRDPEAGAVRLGIFPTLGPYLLPHVVPSIRYRFPQLELLLVEEKSDELLAQLREGK
LDAALLALPLHDEQLHTEFLFEEPFVLAVPEGHPLATRREMTMEELADERLLLLQDGHCLREQALDVCHM
TGASEKSEFQATSLETLRQMVVANVGITLLPLLSVKPPVVCSESIRLINFPLDKQPSRRIAMVWRRSSAM
TTFLERFSSMFKELPKELFDLPQTVVLYKGR

>gi|28198653|ref|NP_778967.1| oxidative stress transcriptional regulator [Xylella fastidiosa Temecula1]
MNLRDLKYLIALADYKHFGRAATACFVSQPTLSTQIKKLEGELGVSLVERAPRKVMMTPAGREAAIRARS
IVAEVEEMKEAARRSRDPEAGAVRLGIFPTLGPYLLPHVVPSIRYRFPQLELLLVEEKSDELLAQLREGK
LDAALLALPLHDEQLHTEFLFEEPFVLAVPEGHPLATRREMTMEELADERLLLLQDGHCLREQALDVCHM
TGASEKSEFQATSLETLRQMVVANVGITLLPLLSVKPPVVCSESIRLINFPLDKQPSRRIAMVWRRSSAM
TTFLERFSSMFKELPKELFDLPQTAVLYKGR

>gi|58583270|ref|YP_202286.1| oxidative stress transcriptional regulator [Xanthomonas oryzae pv. oryzae KACC10331]
MNLRDLKYLVALADHKHFGRAATACFVSQPTLSTQIKKLEDELGVSLVERAPRKVMLTPAGREAALRARS
IVAEVEQMKEAARRSQDPEAGTVRLGIFPTLAPYLLPHVVPRIRERFPRLELLLIEEKSDQLMHQLREGR
LDAALLALPLQDDQLHAEFLFEEPFVLAVPEGHPLSRHDNMTLDALSEQRLLLLGDGHCLREQALDVCHL
AGALEKSEFQATSLETLRQMVAANVGVTLLPLLAVKPPVARSDNIRLIRFRDDKQPSRRIGMAWRRSSAM
TAFLDQLSQLFKELPDSLFTLDQPAAGPKAVAA

>gi|54297727|ref|YP_124096.1| Hydrogen peroxide-inducible genes activator [Legionella pneumophila str. Paris]
MNLRDLHYFVILADVKHFGEAAKRCFVSQPTLSMQIKKLEEELGVVLFERTNKQVLLTDQGSKLLDRTRK
ILILIDEMKELARQSEDPFTGELRLGVIPTVSPYMLPLVMPELKNEYPRLKVWLIEDKTHRLITKLEQGE
LDVAIMALPIDKRFSCQILYEEKFYFACANTHPLAQAKSVNINDLKNQPIMLLEEGHCLREQAMAVCQLA
KADDIADFTATSLETLRLMVQAGMGVTLLPALSTLTASTNHLKCIPFSEPAPSRILGLFWRAGTPRQVCF
NAVAELITKNVQSKLA

>gi|84625106|ref|YP_452478.1| oxidative stress transcriptional regulator [Xanthomonas oryzae pv. oryzae MAFF 311018]
MNLRDLKYLVALADHKHFGRAATACFVSQPTLSTQIKKLEDELGVSLVERAPRKVMLTPAGREAALRARS
IVAEVEQMKEAARRSQDPEAGTVRLGIFPTLAPYLLPHVVPRIRKRFPRLELLLIEEKSDQLMHQLREGR
LDAALLALPLQDDQLHAEFLFEEPFVLAVPEGHPLSRHDNMTLDALSEQRLLLLGDGHCLREQALDVCHL
AGALEKSEFQATSLETLRQMVAANVGVTLLPLLAVKPPVARSDNIRLIRFRDDKQPSRRIGMAWRRSSAM
TAFLDQLSQLFKELPDSLFTLDQPAAGPKAVAA

>gi|71900667|ref|ZP_00682791.1| regulatory protein, LysR:LysR, substrate-binding [Xylella fastidiosa Ann-1]
MNLRDLKYLIALADYKHFGRAATACFVSQPTLSTQIKKLEGELGVSLVERAPRKVMMTPAGREAAIRARS
IVAEVEEMKEAARRSRDPEAGAVRLGIFPTLGPYLLPHVVPSIRYRFPQLELLLVEEKSDELLAQLREGK
LDAALLALPLHDEQLHTEFLFEEPFVLAVPEWHPLATRREMTMEELADERLLLLQDGHCLREQALDVCHM
TGASEKSEFQATSLETLRQMVVANVGITLLPLLSVKPPVVCSESIRLINFPLDKQPSRRIAMVWRRSSAM
TTFLERFSSMFKELPKELFDLPQTVLMYKGR

>gi|94490867|ref|ZP_01298093.1| hypothetical protein CburD_01002045 [Coxiella burnetii Dugway 7E9-12]

MNIRDLKYLLAVADSAHFGKAAEKCFVSQPTLSAQLKKLEEELGVRLFERNNKRVLITPIGQIIAAQVRV
ILQEVEKLKVLAQNAQDPFAGVFHLGIIPTLGPYLLPIILEIFKKRLPKLNLVVYENKTENILHELQQGR
LDAVILALPVSAPNLVVQELFCEPFYVALPKHHPLAKKKSVTLADLEKETLLLLEEGHCLREQALEACSM
TAAKTETGFKATSLETLRHLVAAGAGITLLPALSVNAEKSELAIKSFNATIPSRSIGMLWRDFSARKECC
ETMAKLISAEVKKHPKLKTRAPLKVMERKLE

>gi|29654767|ref|NP_820459.1| hydrogen peroxide-inducible genes
activator OxyR [Coxiella burnetii RSA 493]
MNIRDLKYLLAVADSAHFGKAAEKCFVSQPTLSAQLKKLEEELGVRLFERNNKRVLITPIGQIIAAQVRV
ILQEVEKLKVLAQNAQDPFAGVFHLGIIPTLGPYLLPIIFEIFKKRLPKLNLVVYENKTENILHELQQGR
LDAVILALPVSAPNLVVQELFCEPFYVALPKHHPLAKKKSVTLADLEKETLLLLEEGHCLREQALEACSM
TAAKTETGFKATSLETLRHLVAAGAGITLLPALSVNAEKSELAIKSFNATIPSRSIGMLWRDFSARKECC
ETMAKLISAEVKKHPKLKTRAPLKVMERKLE

>gi|89357996|ref|ZP_01195818.1| oxidative stress transcriptional
regulator [Xanthobacter autotrophicus Py2]
MTLRELQYLVALADHRNFRRAAEACLVSQPTLSTQLRKLEEELGVPLVERAPRRVMLTPAGREAVERARR
ILDEVEQLKEGARRSCAAEAGALKLGVFPTLGPYLLPHVVPLIRARFPELELLLFEEKSAALISRLNYGT
LDAAFLALPVHDSHFHAEFLFEEPFLLAVPGTHALASRDNLSITELSRYNLMLLEDGHCLRDQALDVCQM
AGAREKSEFRATSLETLRQMVAAGVGMTLLPMLATRTPSQPAENIHLLEFSDSKPSRQIAMLWRKTSAMG
RLLADVAQVCRTLPQELLAPRH

>gi|77359710|ref|YP_339285.1| hydrogen peroxide-inducible genes
activator similar to OxyR (but also to other LysR like activators)
[Pseudoalteromonas haloplanktis TAC125]
MNMNLKDLEYVKAIAHFKHFRKAADACFVSQPTLSGQVKKLEQELGVTLFDRSTKQVTLTAKGERLLTQI
NVILEQTQILKELAATSNEPLQGKLTIGIIPTIAPYLLPVLLTSMKEAFINSRFSFIEMQTATILEALDN
GELDFAILADVPELKKYHSVNLYKEDFLVAVSHDNSLAQQKKVALRELQGCSLLMLSDGHCFKDQAQQFC
FSAGVNVSSQYQGNSLETLLALVAMDDGVTFVPKLACTERVGVNYLAIYPNQQRNIVFACRKHYPHLSGV
EQLGEWLSAHPNLKAKLTKSLN

>gi|88811285|ref|ZP_01126541.1| oxidative stress transcriptional
regulator [Nitrococcus mobilis Nb-231]
MVQINLRDLRYLVAVANHRHFGRAAAACYVSQPTLSTQLKKLEQQLGVQLIERNSKQVMLTQAGKMIAER
AHRVLNEVADIVDAARAAGDPMAGDLRLGLIPTVGPYLLPHLIPVLRDVCPRLKPLLYEEQTRALVTRLH
RGELDAALMAVPVNDPRLHFTSLFHEPFYLALPAEHWLARGQHIELGDLEGEHILLLEEGHCLRDQALDV
CDLAGASDIAEFHATSLETLRQMVALGAGVTLLPALAAAANAAVPNHAAIELRPFQQPVPQREMALYWRK
GAAREPALHALADLIRNLSVVRALREPKQANHSAA

>gi|120575115|gb|EAX31739.1| hydrogen peroxide-inducible genes
activator [Coxiella burnetii 'MSU Goat Q177']
MNIRDLKYLLAVADSAHFGKAAEKCFVSQPTLSAQLKKLEEELGVRLFERNNKRVLITPIGQIIAAQVRV
ILQEVEKLKVLAQNAQDPFAGVFHLGIIPTLGPYLLPIILEIFKKRLPKLNLVVYENKTENILHELQQGR
LDAVILALPVSAPNLVVQELFCEPFYVALPKHHPLAKKKSVTLADLEKETLLLLEEGHCLREQALEACSM
TAAKTETGFKAKATSLETLRHLVAAGAGITLLPALSVNAEKSELAIKSFNATIPSRSIGMLWRDFSARKE
CCETMAKLISAEVKKHPKLKTRAPLKVMERKLE

>gi|94494050|ref|ZP_01301253.1| hypothetical protein Rgryl_01000580
[Rickettsiella grylli]
MNFRDLSYLLALAEYRHFGKAAKACSVSQPTLSIQLKKLEQTLGMKLFERGQKKVLMTTSGLRMVEKAKH
IVQAVDEFKRFAKLEKDPFLAELRLGVISSLCPYLLPYILPSIMQELPKITLYLYEDKTENLLIQLKEGK
LDAVVLALPIPHKGLYLRPLFKEPFFLIMPRSHALYDAKKLDLNDLGHYNLLLLEEGHCFRDQALDVCHK
RSNLKEKTNYRATSLETLRHMVGTGAGITLLPLLALETHPFIKNVPLASPVPERKIGMLWRKGSALERCC
KKIATLIENNIPNVITHLEKKLQSKHRMHR

>gi|74318341|ref|YP_316081.1| oxidative stress transcriptional regulator [Thiobacillus denitrificans ATCC 25259]
MTLQELRYLVALADHGHFGRAAEACFITQSTLSTQIKKLEDFLGVTLFDRSLKRVTPTPIGREILQAART
IVDEAERIRTLAKHAQDPMTRTVHLGVIPTLGPYYLPHALTLVHRKHPGLRLLLREEMTPQILEHLADGK
LDAGLLALPVTDEGLRVEPLFHEPFYAALPADHALAAREALSVADIMAEKLLLLDEGHCLREQALDVCGA
RSSGREEVRATSLETLRQMVGMGLGLTLLPALAVDAAPRQTRKLVEIRPFRSPPPGRTIGLVWRRRAPFP
ETFERLAATLKASLPAGVEAV

>gi|121999203|ref|YP_001003990.1| transcriptional regulator, LysR family [Halorhodospira halophila SL1]
MNLRDLRYLVAVAEHRHFGRAARACYVSQPTLSTQLKKLEEYLEVQLVERNRRRVLLTPLGERLAERARS
ILSAVDDMVEVARAQAEPMTGDVRLGVIPTAGPYLLPHVIPDLAQSYPRLRLHLREDLTQRLLDQLRAGS
LDGAILASPIAGDDLVSEPLCHEPFYLAVPRGHDLDRPEPVDAKDLQQTELMLLEEGHCLREQALELCRR
NDVGEAAAFRATSLETLRQMVAAGVGVTLLPALAAAASRLGPDHAAISLRPFAEPAPSRDLALYWRVGTA
REPTFRELVERMRSAAVLQDPTQTLPAA

>gi|115422998|emb|CAJ49528.1| hydrogen peroxide-inducible genes activator [Bordetella avium 197N]
MTLTELKYIVAVARERHFGRAAEACFVSQPTLSVAIRKLEDELGVTLFERGGSEVGVTPIGQRIVAQAQK
VLEESASIKEIARQGHDPLAGPLRVGVIHTIGPYLLPRLVPLQIERTPQMPLLLQENFTVRLVELLRQGE
IDCAIMALPLPEAGLVTQPLYDEPFLVAVPNDHEWAQRQSIDAQDLKQQTMLLLGSGHCFRDQVLEVCPE
LSRFAATSDGIQRTFEGSSLETIRHMVAAGIGVTVLPVTAVPEQATSKSLITYVPFEGAAPTRRVVLAWR
RSFPRMAAVEALAQAVYACGLPGVTMLDDEAAQSQGESLLA

>gi|149928262|ref|ZP_01916505.1| oxidative stress-inducible genes activator [Limnobacter sp. MED105]
MTLTELKYIVALAREKHFGRAADACFVSQPTLSVAIKKLEEELSVSLFERGSNEVSLTPVGERIVVQAQR
VLEEASAIKSIAQQGMDPLAGPLRVGVIYTIGPYLLPGLVSSMIERVPSMPLVLQENFTVRLLELLKQGE
IDVAVLALPINESGFVIQPLYDEPFMVALPKSHRWAHEKTINSDDLRSENMLLLGTGHCFRDQVLGVCPE
LSRFSQSSEGIQRTFEGSSLETIRHMVASGVGITVLPSSSVPNPVPKESLLTYIPLADDDTRRTVALVWR
KSFGRREALEALRDAIMECDLNGVEFLDAPQMVR

>gi|33592699|ref|NP_880343.1| probable LysR-family transcriptional regulator [Bordetella pertussis Tohama I]
MTLTELKYIVAVARERHFGRAAEACFVSQPTLSVAIRKLEDELGVTLFERGGTEVGVTPIGQRIVAQAQK
VLEESASIKEIARQGHDPLAGPLRVGVIHTIGPYLLPRLVPEQIARTPQMPLLLQENFTVRLVELLRQGE
IDCAIMALPLPEAGLVMQPLYDEPFVVAVPHDHEWAQRKAIDAQDLKQQTMLLLGSGHCFRDQVLEVCPE
LSRFSASSDGIQRTFEGSSLETIRHMVAAGIGVTVLPFTAVPNPPQPKSLLRYLPFDGETPERRVVLAWR
RSFPRLAAIEALAQAVYACGLPGVRMLDEEAASAQVD

>gi|104784350|ref|YP_610848.1| transcriptional regulator, LysR family [Pseudomonas entomophila L48]
MTLTELRYIVTLAQEQHFGHAAERCHVSQPTLSVGVKKLEDELGVLIFERSKSAVRLTPVGESIVAQAQK
VLEQAQGIRELAQAGKNQLTAPLKVGAIYTVGPYLFPHLIPQLHRVAPQMPLYIEENFTHVLREKLRNGE
LDAVIIALPFNEADVLTLPLYDEPFCALMPADHPWTQKDTIDTAMLNDKSLLLLGEGHCFRDQVLEACPT
LNKGGEGSKHTTVESSSLETIRHMVASGLGVSILPLSAVHSHHYAPGVIEVRPLTAPAPFRTVAIAWRAS
FPRPKAIEILADSIRLCSVAKAPVEQPA

>gi|66043472|ref|YP_233313.1| regulatory protein, LysR:LysR, substrate-binding [Pseudomonas syringae pv. syringae B728a]
MTLTELRYIVTLAQEQHFGHAAERCHVSQPTLSVGVKKLEDELGVLIFERSKSAVRLTPVGEGIVAQAQK
VLEQAQGIRELAQAGKNQLTAPLKVGAIYTVGPYLFPHLIPQLHRVAPQMPLYIEENFTHVLRDKLRNGE
LDAVIIALPFNEADVLTLPLYDEPFSVLMPADHPWTQKETIDASALNDKSLLLLGEGHCFRDQVLEACPT

LGKGNEGAKHTTVESSSLETIRHMVASGLGISILPLSAVDSHHYAPGVIEVRPLTPPVPFRTVAIAWRAS
FPRPKAIEILADSARLCSVARPKTVAS

>gi|126356271|ref|ZP_01713276.1| transcriptional regulator, LysR family
[Pseudomonas putida GB-1]
MTLTELRYIVTLAQEQHFGHAAERCHVSQPTLSVGVKKLEDELGVLIFERSKSAVRLTPVGESIVAQAQK
VLEQAQGIRELAQAGKNQLTAPLKVGAIYTVGPYLFPHLIPQLHRVAPQMPLYIEENFTHVLREKLRNGE
LDAVIIALPFNEADVLTLPLYDEPFCALMPADHPWTAKKTIDTAMLNDKSLLLLGEGHCFRDQVLEACPT
LNKGGEGSRHTTVESSSLETIRHMVASGLGVSILPLSAVHSHHYAPGVIEVRPLTAPAPFRTVAIAWRAS
FPRPKAIEILADSIRLCSVAKNPAEQPA

>gi|146280532|ref|YP_001170685.1| probable transcriptional regulator
[Pseudomonas stutzeri A1501]
MLTMTLTELRYIVTLAQEQHFGRAAERCHVSQPTLSVGVKKLEDELGVLIFERTKSAVRLTPVGEGIVTQ
AQKVLEQAQSIRELAQVGKNQLAAPLKVGAIYTVGPYMFPHLIPQLHRVAPDMPLYIEENFTHVLRDKLR
TGELDAIIIALPFQEADVLTKPLYDEPFYVLMPADHPWTAKETIDAEMLNDKSLLLLGEGHCFRDQVLEA
CPTVRKGEAASHTTVESSSLETIRHMVASGLGVSILPLSAVESHHYSPGVLEIRPLTPPVPFRTVAIAWR
ASFPRPKAIEILADSIRLCSVGKPPSAKA

>gi|26991985|ref|NP_747410.1| transcriptional regulator, LysR family
[Pseudomonas putida KT2440]
MTLTELRYIVTLAQEQHFGHAAERCHVSQPTLSVGVKKLEDELGVLIFERSKSAVRLTPVGENIVAQAQK
VLEQAQGIRELAQAGKNQLTAPLKVGAIYTVGPYLFPHLIPQLHRVAPQMPLYIEENFTHVLREKLRNGE
LDAVIIALPFNEADVLTLPLYDEPFCALMPADHPWTAKKTIDTAMLNDKSLLLLGEGHCFRDQVLEACPT
LNKGGEGSRHTTVESSSLETIRHMVASGLGVSILPLSAVHSHHYAPGVIEVRPLTAPAPFRTVAIAWRAS
FPRPKAIEILADSIRLCSVAKNPAEQPA

VITA

Name:	Gwendowlyn Sue Knapp
Address:	1141 Salt Creek Road Chesterton, IN 46304
Email Address:	gwen.knapp@gmail.com
Education:	B.S., Molecular and Genetic Biology, 1999, Purdue University, West Lafayette, IN Ph.D., Biochemistry, 2008, Texas A&M University, College Station, TX
Professional Publications:	Bryan, B.A., Knapp, G.S., Bowen, L.M. and Polymensis, M. 2004. The UV response in *Saccharomyces cerevisiae* involves the mitogen-activated protein kinase Slt2p. *Current Microbiology* **49**: 32-34. Kopytek, S.J., Dyer, J.C., Knapp, G.S. and Hu, J.C. 2000. Resistance to methotrexate due to AcrAB-dependent export from *Escherichia coli. Antimicrob Agents Chemother.* **44**: 3210-3212.