

LIGAND RECOGNITION BY PROKARYOTIC  
TRANSCRIPTION REGULATORS

by

Joy Lynne Huffman

A DISSERTATION

Presented to the Department of  
Biochemistry and Molecular Biology  
and the Oregon Health Sciences University  
School of Medicine

in partial fulfillment of  
the requirements for the degree of  
Doctor of Philosophy

April 2001

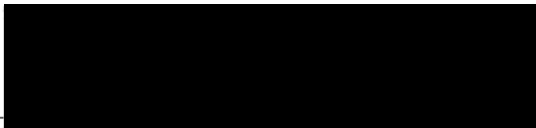
School of Medicine  
Oregon Health Sciences University

---

CERTIFICATE OF APPROVAL

---

This is to certify that the Ph.D. thesis of  
Joy Lynne Huffman  
has been approved

  
Professor in charge of thesis

  
  
Member

  
Member

  
Member

## **TABLE OF CONTENTS**

<b>List of tables and figures</b>	iii
<b>Acknowledgments</b>	vi
<b>Abstract</b>	vii
<b>Chapter 1. Introduction</b>	1
Overview of transcription regulation	1
RNA polymerase	2
Regulation of prokaryotic transcription	3
Repressors and activators	4
The protein-DNA interface	6
DNA binding motifs	10
The helix-turn-helix motif	11
Winged helix proteins	15
The $\beta$ -ribbon motif	21
The <i>E. coli</i> purine repressor	27
The LacI/GalR family	27
Regulation of purine biosynthesis by PurR	29
Structures of PurR	30
DNA recognition	32
Allosteric mechanism	34
Corepressor binding and specificity	35
Prokaryotic heavy metal transcription regulators	38
Metalloregulatory proteins that are not MerR family members	39
The MerR family	42
The BmrR-drug-DNA structure	45
The <i>Proteus mirabilis</i> transcription regulator, PMTR	46
Tables and Figures	50
<b>Chapter 2. Determination of macromolecular structures by X-ray crystallography</b>	85
Crystallization	86
Crystals: symmetry and space groups	90
X-rays: diffraction and the reciprocal lattice	93
The structure factor	96
Assessing the quality of data collection	97
Phase determination	98
Model refinement	100
Assessing the quality of a model	102
Tables and Figures	104

<b>Chapter 3. Structural and functional characterization of PurR allosteric mutants W147F, W147A, and W147R</b>	116
Abstract	117
Introduction	118
Materials and methods	121
Results	127
Discussion	136
Tables and figures	142
 <b>Chapter 4. Structure-based redesign of corepressor specificity of the <i>Escherichia coli</i> purine repressor by substitution of residue 222</b>	150
Abstract	151
Introduction	153
Materials and methods	155
Results	160
Discussion	171
Tables and figures	177
 <b>Chapter 5. Towards redesign of corepressor specificity of the <i>Escherichia coli</i> purine repressor by substitution of residues 190 and 192</b>	188
Abstract	189
Introduction	191
Materials and methods	195
Results	201
Discussion	207
Tables and figures	212
 <b>Chapter 6. Biochemical characterization of MerR family member <i>Proteus mirabilis</i> transcription regulator , or PMTR</b>	233
Abstract	234
Introduction	235
Materials and methods	239
Results	246
Discussion	251
Tables and figures	256
 <b>Chapter 7. Summary and conclusions</b>	280
 <b>Literature cited</b>	287



## List of tables and figures

### **Chapter 1.**

- Table 1.1 A taxonomy of prokaryotic transcription regulators for which DNA binding domain structures are available.
- Table 1.2 Alignment of phage  $\lambda$   $O_R$  and  $O_L$  sites.
- Figure 1.1A. The transcription initiation complex. B. The four steps of transcription initiation by *E. coli* RNA polymerase.
- Figure 1.2A. Watson-Crick base pairs as found in B-DNA. B. Local DNA helix parameters for pairs of adjacent base pairs
- Figure 1.3 Helix-turn-helix motifs from three proteins bound to DNA. A.  $\lambda$  cI. B. Trp repressor. C.  $\lambda$  Cro.
- Figure 1.4A.  $\lambda$  cI bound to  $O_L1$  DNA. B.  $\lambda$  Cro bound to pseudo-consensus DNA.
- Figure 1.5 Close-up view of CRP bound to DNA.
- Figure 1.6 Winged helix domains from four proteins: A. OmpR. B. BirA. C. BmrR. D. MuR.
- Figure 1.7A. The BmrR-DNA-drug complex. B. B-form DNA and DNA bound to BmrR.
- Figure 1.8A. MetJ bound to S-adenosyl methionine and DNA. B. The Arc repressor bound to DNA.
- Figure 1.9 Integration host factor bound to 35 bp of DNA.
- Figure 1.10A. Genes regulated by PurR. B. sequence of *pur* regulon operators recognized by PurR. C. The loci of each PurR operator site.
- Figure 1.11A. The PurR-hypoxanthine-*purF* operator complex. B. The PurR corepressor binding domain.
- Figure 1.12 Illustrations of the effector binding domains of monomers of LacI/GalR family members and a periplasmic binding protein. A. PurR. B. Ribose binding protein. C. LacI. D. Trehalose repressor.
- Figure 1.13 Close-up views of the PurR corepressor binding pocket in the A. apo and B. holo structures.
- Figure 1.14A. The structure of SmtB. B. Two dimers of DtxR bound to DNA
- Figure 1.15 Cartoon depicting the interactions of PMTR with DNA and RNA polymerase.

### **Chapter 2.**

- Table 2.1 The seven crystal systems.
- Figure 2.1 The phase diagram for crystallization.
- Figure 2.2 A milliliter plate used in hanging drop vapor diffusion experiments. Inset: One well in such an experiment.
- Figure 2.3 The solubility of enolase and carbonmonoxyhemoglobin as a function of ionic strength.
- Figure 2.4 The unit cell.
- Figure 2.5 The fourteen Bravais lattices.
- Figure 2.6A. A four-fold rotation axis. B. A  $3_1$  screw axis. C. A  $3_2$  screw axis.
- Figure 2.7 An illustration of Bragg's law.

Figure 2.8 The Fourier transforms of a set of lines and a lattice.

Figure 2.9 Orthorhombic and monoclinic direct and reciprocal cells.

Figure 2.10 The Ewald construction.

Figure 2.11A. A Ramachandran plot model. B. The Ramachandran plot for wild-type PurR bound to hypoxanthine and the *purF* operator.

### Chapter 3.

Table 3.1 *In vivo* repression and *in vitro* equilibrium binding of wild-type PurR and W147 mutants

Table 3.2 Data collection and refinement statistics for W147F corepressor binding domain and mutant PurR-hypoxanthine-*purF* operator complexes

Figure 3.1 The corepressor binding domain from A. the holo PurR-hypoxanthine-*purF* operator structure, and B. the apo corepressor binding domain structure. Close-up views of the corepressor binding pocket in the: C. PurR-hypoxanthine-*purF* operator complex structure and D. corepressor free PurR CBD structure.

Figure 3.2 Representative binding isotherms for wild-type and W147 mutant proteins.

Figure 3.3 Overlays of the corepressor binding pockets of each of the mutant structures solved onto the wild-type structure.

### Chapter 4.

Table 4.1 Affinities of wild-type and mutant purine repressors for guanine, hypoxanthine and xanthine.

Table 4.2 Summary of selected crystallographic data collection and refinement statistics for E222A, Q, and K mutants bound to hypoxanthine (hyp), guanine (gua), or xanthine (xan) and *purF* operator.

Figure 4.1 A close-up view of the corepressor binding pockets of the wild-type PurR-hypoxanthine-*purF* operator complex, in A, and PurR-guanine-*purF* operator complex, in B.

Figure 4.2 The electrostatic surfaces and partial charges of: A. guanine, B. hypoxanthine, and C. xanthine.

Figure 4.3 Views of the corepressor binding pocket of the mutants after the superimposition of their C $\alpha$  carbon atoms onto those of the wild-type PurR protein bound to hypoxanthine.

Figure 4.4 Electrostatic environment of corepressors.

### Chapter 5.

Table 5.1 Affinities of wild type and mutant purine repressors for *purF* operator alone or in the presence of hypoxanthine or ribose.

Table 5.2 Summary of selected crystallographic and refinement data for R190K-hypoxanthine-*purF* operator complex.

Figure 5.1A. The LacI/GalR superfamily phylogenetic tree. B. Primary sequence alignment of PurR and RbsR.

Figure 5.2 Close-up views of the ligand binding sites of A. PurR and B. RBP.

Figure 5.3 Equilibrium binding isotherms for titrations of wild-type and mutant proteins into fluoresceinated *purF* operator and hypoxanthine, ribose, or no corepressor.

- Figure 5.4 Equilibrium binding isotherms for titrations of hypoxanthine into fluoresceinated *purF* operator and wild-type or mutant proteins.
- Figure 5.5 Equilibrium binding isotherms in corepressor-independent Buffer System I.
- Figure 5.6A. Percent DNA binding plotted against ribose concentration. B. Change in polarization (mP) upon addition of protein plotted against concentration of ribose.
- Figure 5.7 View of the corepressor binding pocket of the R190K mutant after the superimposition of their C $\alpha$  carbon atoms onto those of the wild type PurR protein bound to hypoxanthine.
- Figure 5.8 Model of the Thr192Pro mutation, based on the wild-type PurR-hypoxanthine-*purF* operator structure.

## Chapter 6.

- Table 6.1 Percent identity between operators bound by select MerR family members.
- Table 6.2 Percent identity between select MerR family members, in total, in the DNA binding domain alone, and in the linker/effector binding domain.
- Figure 6.1A. Schematic of the *P. mirabilis* genomic fragment inserted into the PROT1 plasmid. B. The intergenic regions used in EMSA and footprinting experiments.
- Figure 6.2 Electrophoretic mobility shift assay results for A. *atpaseP* and B. *pmtrP*.
- Figure 6.3 Footprinting results for: A. *atpaseP*-F and B. *atpaseP*-R DNA.
- Figure 6.4 Linear traces of radioactivity down sequencing gel lanes from hydroxyl radical DNA cleavage reactions. A. Template strand. B. Non-template strand.
- Figure 6.5 Summary of footprinting results. A. Results in the absence of zinc. B. Results in the presence of 250  $\mu$ M ZnCl<sub>2</sub>.
- Figure 6.6 Fluorescence polarization equilibrium binding isotherms for PMTR.
- Figure 6.7 CPM labelling experiment results.
- Figure 6.8 Sequence alignments. A. The operator identified in our PMTR footprinting experiments aligned with operators bound by other MerR family members. B. PMTR and select other MerR family members.
- Figure 6.9 Model of the dimeric PMTR-DNA complex.
- Figure 6.10 Close-up view of the model of one PMTR DNA binding domain monomer bound to one DNA half-site.

## Acknowledgments

I am deeply indebted to many people, without whom this dissertation would not have been possible. First, I must thank my amazing advisor, Dick Brennan, around whom things always "turn out fine", for encouraging me, reading my horrid first drafts of things, helping me figure "stuff" out, providing an easy-going lab atmosphere, and for actually caring about what I was doing and when. I wish to thank Katya Heldwein for giving me inspiration when I needed it most, for helpful comments throughout my graduate career, and the occasional, "I know, it sucks. Just keep doing it." I also want to thank Maria Schumacher, for being helpful even if she didn't mean to be and setting an ineffable example of what was possible. Many thanks go to Sam Rigby, for putting up with my thesis-writing hysteria, for thinking it's "cool" that I'm a biochemist, for critically reading large portions of this tome, and for being himself. I should also thank Samantha, my cat, for sitting on my lap during most of the writing of this thesis and not throwing up too often.

I wish to dedicate this dissertation to my family, in particular, to all of my parents. To my mom, who died of cancer when I was 16. Her influence is still strong and always will be. She was generous and kind, and always lots of fun. To my dad, who has quietly put up with all of my changing styles, loud music, and crazy ideas. He may think I'm a little nuts, but hopefully he'll believe I've turned out okay anyway. To my birthmom, who is a wonderful person I am looking forward to getting to know in years to come.

## Abstract

This thesis examines two aspects of transcription regulation in prokaryotes. The first focuses on corepressor binding and the allosteric switch of the *Escherichia coli* purine repressor, PurR, a member of the LacI/GalR family. The second is the biochemical characterization of the *Proteus mirabilis* transcription regulator, PMTR, a MerR family member. Diverse approaches and tools are used in each study, primarily structural in the former and biochemical in the latter.

PurR regulates the transcription of genes involved in *de novo* purine nucleotide biosynthetic and salvage pathways by repressing the transcription of that regulon when bound by a corepressor, either hypoxanthine or guanine. A great deal of biochemical and genetic data is available on PurR, including the structures of various PurR-corepressor-DNA complexes and the apo structure. Using these data, additional experiments tested the functional importance of residues Trp147, Glu222, Arg190, and Thr192 were done.

The role of the indole ring of Trp147, a key allosteric switch residue, was tested by its substitution with alanine, arginine, or phenylalanine. *In vitro* ligand binding studies and crystal structures are presented in Chapter 3. The latter reveal the basis for the altered corepressor affinities, which arises from the conformational preference of each mutant. Trp147Ala/Arg prefer the DNA-bound conformation, whilst Trp147Phe prefers the apo. The importance of Trp147 in the allosteric switch is revealed in the altered structural stability of each mutant in the holo and apo conformations.

To understand better the corepressor specificity of PurR, and to test that understanding by attempting to expand the corepressor repertoire to include xanthine, site-directed mutagenesis, corepressor binding, and structural studies were done. In

Chapter 4, *in vivo* expression data, *in vitro* corepressor binding data, and X-ray structures are presented for Glu222 to alanine, glutamine, and lysine substitutions in complex with hypoxanthine, guanine, or xanthine. As predicted, these studies reveal an increase in mutant:xanthine affinity, with the Glu222Lys protein binding all three corepressors with physiologically relevant affinities.

In Chapter 5, we attempted to engineer a ribose-binding PurR with the ultimate goal of switching the identity of the *in vivo* corepressor from a purine to ribose.

"Corepressor" and DNA binding data for three RbsR-like PurR proteins, in which PurR residues Arg190 and Thr192 were replaced by lysine and proline (the complementary RbsR residues), singly and in combination, were obtained. This simple strategy failed, but suggested other avenues to pursue.

A second focus of this thesis was the biochemical characterization of PMTR, the first detailed analysis of any *P. mirabilis* transcription factor. In this work, a PMTR cognate DNA-binding site was identified and characterized using DNA cleavage protection, electrophoretic mobility shift, and fluorescence polarization assays. The relative metal binding affinity of PMTR *in vitro* was also determined and appears to be different from other MerR family members in that PMTR binds zinc, cadmium, and copper. Thus, PMTR is a zinc-binding MerR family member with strong affinity for two additional toxic metals.

## Chapter 1. Introduction

### Overview of transcription regulation

Proper regulation of transcription is crucial for all living cells, for it is at this level that gene expression is mainly controlled. Gene expression must be sensitive to environmental signals—genes must be “turned on” in response to some signals and “turned off” in response to others. Any change in a cell’s environment triggers the alteration of the expression of at least one gene. Almost every process of life is controlled by gene regulation at some level: catabolic or anabolic enzymes must be made at times when building blocks such as amino acids or nucleotides are required or in excess, toxic substances must be extruded from cells by multidrug transporters, and membrane proteins functioning as transporters, channels, or pumps must be expressed in proper proportions depending on the needs of the cell. Simply transcribing all genes all of the time would be energetically very costly and an inefficient use of resources. Furthermore, this would not allow the development or differentiation of multicellular organisms. In all organisms, improperly regulated transcription—with genes being always “on” or always “off”—usually leads to sickness and eventually death.

Jacob and Monod first put forth the idea that gene expression is controlled by binding of a particular protein to a regulatory region, or promoter, based on experiments on sugar metabolism in *E. coli* (Jacob and Monod, 1961). Since those early days, several prokaryotic transcription regulatory systems have been extensively characterized at the molecular and atomic levels. We are now poised to delve deeply into mechanistic questions regarding specific transcription regulators.

In this dissertation, two aspects of transcription regulation will be discussed. The first involves the *Escherichia coli* purine repressor, PurR, a member of the LacI family, and deals with effector binding. The goal of these studies was to understand the chemical features of PurR that contribute to its specificity for its physiological corepressors, hypoxanthine and guanine, and to understand how the binding of corepressor induces the transcription factor to change the regulation of a gene. To address these questions, studies were carried out on the *E. coli* purine repressor, a particularly well characterized system suitable for experiments regarding structure-function relationships. The crystal structures of apo PurR and PurR bound to hypoxanthine and cognate DNA, along with a wealth of biochemical and mutagenesis data, have enabled us to study the effects of individual residues on the structural and functional contributions of individual residues.

The second focus of this thesis is on biochemical characterization of the *Proteus mirabilis* transcription regulator, PMTR. This protein belongs to a different family of prokaryotic transcription regulators, the MerR family. The experiments reported here utilize an array of molecular biological and biochemical techniques different from those used with PurR. These studies describe the molecular and thermodynamic parameters of DNA binding by this protein and allow us to make comparisons to related proteins.

### RNA polymerase

Transcription of prokaryotic genes is accomplished by the large, multisubunit enzyme, RNA polymerase (RNAP). *E. coli* RNAP consists of five subunits that assemble on DNA in the order  $\alpha$ ,  $\alpha_2$ ,  $\alpha_2\beta$ ,  $\alpha_2\beta\beta'$ ,  $\alpha_2\beta\beta'\sigma$  (Busby and Ebright, 1994).  $\alpha_2\beta\beta'$  is termed the core polymerase, and when joined with  $\sigma$  it is called the holoenzyme,



illustrated bound to DNA in Figure 1.1A.  $\sigma^{70}$ , named for its molecular weight of 70 kDa, is the most commonly found  $\sigma$  protein, but other, more specialized  $\sigma$  proteins exist.  $\alpha$  initially binds to the UP (upstream) element of the promoter, if one is present, using its C-terminal domain ( $\alpha$ CTD) to bind DNA.  $\alpha$ CTD is a major target for activator proteins, with multiple interaction surfaces mapped for different proteins (Liu and Hanna, 1995; Savery et al., 1998; Zhou et al., 1993). The  $\alpha$ CTD is joined by a long, flexible linker to the N-terminal domain ( $\alpha$ NTD), which contacts the other subunits of RNAP.  $\alpha$ NTD—and to a lesser extent  $\beta$  and  $\beta'$ —have been observed as regulatory targets in only a few cases to date (Lee and Hoover, 1995; Miller et al., 1997; Niu et al., 1994).  $\beta$  and  $\beta'$  are responsible for the catalytic activity of RNAP. The other subunit of RNAP that interacts directly with DNA is  $\sigma^{70}$ , which enables RNAP to recognize promoters. The holoenzyme slides along the DNA until the  $\sigma$  subunit recognizes the -35 and -10 elements of a promoter. Furthermore,  $\sigma^{70}$  is a common target for regulatory protein interactions. For example, phage  $\lambda$  repressor protein, cI, represses host transcription by interacting with the  $\sigma^{70}$  region that contacts the -10 element of DNA (Li et al., 1994).

### Regulation of prokaryotic transcription

Transcription can be divided into three phases: initiation, elongation, and termination. While each is subject to regulatory control, I shall focus on the initiation phase in this work. Transcription initiation is the most highly regulated step in the expression of specific genes, and regulation at the earliest step in transcription provides maximal conservation of energy in the cell.

The initiation of transcription can be broken further into four sequential events, depicted in Figure 1.1B (Herschbach and Johnson, 1993). First, the RNAP holoenzyme must bind to the promoter region of a particular gene, usually located immediately upstream of the transcription start site. Second, the RNAP-DNA complex isomerizes from the “closed” to the “open” form. This step includes the local melting of the DNA to allow access to the single stranded template. Third, the first few phosphodiester bonds of the RNA transcript are formed. Finally, RNAP clears the promoter and the  $\sigma$  factor is released from the complex.

As shown in Figure 1.1A, prokaryotic promoters are simple, consisting of three elements: the -35 region, the -10 region, and the UP element (Busby and Ebright, 1994). The -35 region has the consensus sequence TTGACA, while the -10 has consensus TATAAT. The UP element is an AT-rich region of about 20 base-pairs (bp) located just upstream of the -35 element (Ross et al., 1993), although it is not found in all promoters. Promoters vary in efficiency—meaning their ability to promote transcription of genes immediately downstream—depending on their similarity to these derived consensus sequences. The more similar a given promoter is to the consensus sequences, the higher the level of basal transcription.

#### Repressors and activators

The “ground state” transcription level in *E. coli* favors transcription from most promoters. Thus, prokaryotic transcription is “nonrestrictive”, requiring no additional activators for efficient transcription (Struhl, 1999). Repressor proteins typically regulate promoters that have high sequence similarity to the consensus preferred by RNAP.

Prokaryotic repressor proteins generally function by binding and sequestering -10 or -35 elements, thereby blocking the activity of RNAP. For example, the *E. coli* purine repressor represses transcription of the *purF* gene by binding a site covering DNA -37 to -21 relative to the *purF* transcription start site. About half of all *E. coli* promoters are regulated by activating transcription factors that overcome defects in the -35 or -10 promoter elements (Gralla, 1996). Promoters must be weak or specifically repressed for genes to be regulated by activators.

Activation and repression have been worked out in considerable detail for the *E. coli* cyclic AMP receptor protein, CRP (also known as CAP for catabolite activator protein) and the cytidine repressor, CytR. CRP activates transcription at over one hundred promoters in *E. coli*, binding three different classes of promoters (Class I, II, and III) (reviewed in (Busby and Ebright, 1999)).

While CRP positively regulates the genes involved in the uptake and catabolism of deoxyribonucleosides, CytR and DeoR repress the expression of those genes (Mironov et al., 1989). CytR is a member of the LacI/GalR family of transcription regulators, with significant homology to the purine repressor, PurR (Schumacher et al., 1993). CytR represses transcription of these genes in the absence of its small molecule effectors, cytidine or adenosine. To repress, CytR typically binds DNA between two CRP proteins, thus acting as an "antiactivator" by masking the activation domains of CRP, although only one CRP dimer is required for repression by a CytR dimer (Meibom et al., 2000). CytR also binds close to or overlaps with the DNA binding site of RNAP. Several lines of evidence point to the repression of transcription being mediated by direct protein-protein contacts between CRP and CytR, and, interestingly, not by direct protein-DNA

contacts. These include: (1) Certain mutations of CRP can abolish regulation by CytR; (2) CytR, in which the DNA binding domain has been removed, is still capable of repression in a CRP-dependent manner *in vitro*; and (3) cytidine or adenosine interrupts the interaction between CRP and CytR, not between CytR and DNA (Busby and Ebright, 1999; Valentin-Hansen et al., 1996). This elegant interplay of activating and repressing factors is reminiscent of the complexity of eukaryotic systems and supports the relevance of the study of prokaryotic mechanisms to higher forms of life.

### The protein-DNA interface

There are two major methods of comparing protein-DNA interactions. The most obvious is the characterization of DNA binding motifs based on sequence and fold similarities of the proteins under review, or of the cognate DNA sequences to which they bind. This sort of taxonomy will be used later in this introduction. The other method, although less common and somewhat less specific, can yield valuable information regarding protein-DNA recognition. In this method, large sets of protein-DNA interfaces, known from X-ray crystallographic and nuclear magnetic resonance (NMR) studies, are analyzed for overall themes of structural features, regardless of which organisms the proteins are from or which specific motifs they use to bind DNA. The data generated are then compared to features observed in protein-protein interfaces. A few such general surveys have been carried out recently (Jones et al., 1999; Nadassy et al., 1999), the results of which I shall summarize briefly. Because less than one hundred unique protein-DNA structures have been solved beyond 3.0 Å resolution to date (19 from prokaryotic transcription regulators, Table 1.1), interfaces are not broken into the

kingdoms in which they occur, nor into repressors versus activators, or other categories. These are general analyses based on all double-stranded DNA-protein complexes of high quality. For analyses of hydrogen bonding patterns and water content, only structures of 2.4 Å resolution and above were utilized, of which there were 65 or fewer used in each study.

Nadassy and coworkers found that protein-DNA interactions buried 1600 +/- 400 Å<sup>2</sup> when broken into "DNA recognition modules", which are one protein DNA binding motif and the DNA it contacts (Nadassy et al., 1999). These modules usually combined 12 +/- 3 nucleotides with 24 +/- 6 amino acid residues. Enzymes, which typically envelop DNA rather than using 1 or 2 "reading heads" on average buried an additional 600 Å<sup>2</sup>. This compares well with protein-protein interfaces, which typically bury ~1700 Å<sup>2</sup> in permanent complexes (Lo Conte et al., 1999). The actual numbers vary between studies, but the trends remain the same: the same minimum, standard-size interface is required to form a stable complex, whether it be between a protein and DNA or between two proteins.

When the components of the average surface of an oligomeric protein are compared to the average surface of a protein in contact with DNA, the DNA-binding surface proves to be made up of fewer nonpolar residues, slightly more neutral polar residues, and even more charged residues. In non-DNA binding proteins, the positive and negative charges on the surface are nearly equivalent. In contrast, the charged residues are mostly positive in DNA binding proteins, which is required to counteract the negatively charged phosphate backbone of the DNA. In fact, the most common interactions between proteins and nucleic acids are from protein arginine and lysine

residues to DNA phosphate backbone oxygens. The DNA side of this interaction has almost twice the amount of charged surface area as the average protein at their interfaces.

On average, there are 22 hydrogen bonds in a protein-double-stranded DNA complex. When broken down into recognition modules, the interface has about 13 hydrogen bonds, due to the tendency of proteins to bind DNA as dimers. 90% of these hydrogen bonds are donated from the protein to the DNA, and 60% are from lysine or arginine or the main chain amine group. Additionally, there are about 21 water molecules found at the interface that hydrogen bond to both the protein and the DNA. This is about the same number as there are hydrogen bonds between the two macromolecules, revealing the importance of water in the average protein-DNA interface. Of course, one must keep in mind that these are averages, and examples can be found at any extreme. Tryptophan repressor-DNA contacts are almost exclusively water-mediated (Otwinowski et al., 1988), while the tetracycline repressor, TetR, excludes water from the center of its interface with operator DNA (Orth et al., 2000).

Within the DNA bases, certain recognition elements are used more frequently than others. Polar interactions in the major groove are generally restricted to the 6 and 7 positions of the purine and the pyrimidine 4 position ("W2", Figure 1.2A), which together account for 80% of the hydrogen bonds to bases. Minor groove interactions, while far less common, are typically made with the purine 2 and 3 and pyrimidine O2 exocyclic positions ("S2", Figure 1.2A). About half of all bonds to bases are to guanine, with arginine and lysine responsible for the majority of them. Adenine is usually specifically recognized by the carboxamide group of asparagine or glutamine, with its N6 acting as a hydrogen bond donor and N7 as an acceptor. The cytosine N4 amino group is the most

common hydrogen bond donor in DNA, usually donating to protein carboxylate side chains or main chain carbonyls. The 5-methyl group of thymine is also contacted often and is a major source of sequence discrimination by nonpolar and CH $\cdots$ O bonds (Mandel-Gutfreund et al., 1995).

It must be noted that this method of characterizing the protein-DNA interface does not consider the contributions of direct versus indirect readout. Direct readout is the specific recognition of a particular DNA sequence by direct or water-mediated contacts to the bases, while indirect readout relies on the intrinsic shape of the DNA for specific recognition (Otwinowski et al., 1988).

Proteins can undergo several types of structural changes upon binding DNA. Side chain rotations occur as a consequence of nearly every protein-DNA binding event. Disorder to order transitions are common, but not ubiquitous, in DNA binding domains. For example, PurR undergoes a significant coil-to-helix transition in its DNA binding domain (Nagadoi et al., 1995; Schumacher et al., 1995), yet phage  $\lambda$  Cro repressor shows no change in the amount of ordered structure upon binding DNA (Albright and Matthews, 1998a; Anderson et al., 1981). Proteins can alter their quaternary structure, as in DNA-induced dimerization, or they can move subunits relative to one another. Rigid body domain movements are observed, as are changes in tertiary structure. Cro subunits rotate  $\sim 55^\circ$  with respect to each other (Brennan et al., 1990), as do the N-subdomains of the PurR corepressor binding domain (Schumacher et al., 1995).

Of course, DNA can change structure as well. Proteins can bend DNA either toward the major groove, thus compressing the groove, or toward the minor groove, compressing that groove instead. DNA bending can occur as a very localized distortion,

as in the case of sharp kinks such as those observed with CRP (Schultz et al., 1991), PurR (Schumacher et al., 1994a), and LacI (Lewis et al., 1996), which often result from large roll angles between base pair steps (Figure 1.2B). Small, local distortions that act cumulatively can cause a smooth bend, as seen with the E2 papillomavirus transcription regulator (Hegde et al., 1992) and *S. cerevisiae* cell-type determinant MATa1/ $\alpha$ 2 homeodomain proteins (Li et al., 1995). DNA bending allows a protein to contact more bases within a smaller area. For example, if the integration host factor, IHF, were to contact linear B-DNA, it would reach about 7 base pairs (bp). By wrapping the DNA around itself with large distortions, it instead contacts about 26 bp (Figure 1.9, ref. (Rice et al., 1996)). As will be discussed later, both PurR and PMTR distort DNA in distinct ways that contribute to their binding of cognate DNA.

### **DNA binding motifs**

An extraordinarily small number of protein folds are used to bind a huge number of DNA sequences, which vary in length (the amount of surface recognized) and composition. Many prokaryotic transcription factors are homodimers that bind to palindromic or pseudopalindromic cognate DNA sites. Perhaps the most thoroughly studied DNA binding motif found in prokaryotes is the helix-turn-helix (HTH) motif (reviewed in (Brennan and Matthews, 1989; Wintjens and Rooman, 1996)). This very common motif is characterized by two helices packed together in a specific orientation and joined by a four-residue turn. A third helix stabilizes the motif as a compact, globular domain. A subfamily of the HTH DNA binding proteins is the winged helix (WH) family, which contains a HTH motif followed by one or two  $\beta$  hairpin wings



(reviewed in (Brennan, 1993) and (Gajiwala and Burley, 2000)). The role of these wings in DNA recognition varies.

Another DNA binding motif found in prokaryotes is the  $\beta$ -ribbon motif.

Structures have been solved for several members of this family, including MetJ (Somers and Phillips, 1992), Arc (Breg et al., 1990; Raumann et al., 1994), Mnt (Burgering et al., 1994), and CopG (Gomis-Rüth et al., 1998). In this group of proteins, one  $\beta$  strand from each monomer makes up an antiparallel two-stranded  $\beta$ -ribbon motif that is inserted into the major groove. Several architectural and global transcription-regulating DNA binding proteins, HU from *B. stearothermophilus*, SPO1 TF-1, and *E. coli* integration host factor (IHF), also use  $\beta$  strands to recognize DNA but use them in the minor groove (Rice et al., 1996; White et al., 1999).

#### The helix-turn-helix (HTH) motif

The HTH motif is the most evolutionarily successful, and therefore most widely used DNA binding motif in prokaryotes. Structures have now been solved for about twenty prokaryotic members of the HTH family, including bacteriophage  $\lambda$  repressor (cI),  $\lambda$  Cro, and *E. coli* CRP (Anderson et al., 1981; McKay and Steitz, 1981; Pabo and Lewis, 1982). Almost every prokaryotic protein that uses a HTH motif to bind DNA functions as a homodimer, with each motif binding one half-site of symmetry-related DNA. One exception is MarA, a member of the AraC transcriptional activator family that regulates multiple antibiotic resistance in *E. coli*. MarA binds its cognate DNA as a monomer using a bipartite HTH motif (Rhee et al., 1998). Another exception is the *E. coli* Rob protein, which is also an AraC family member and related to MarA. Rob also

has a bipartite HTH motif, but uses only one of them to make base-specific contacts to the DNA (Kwon et al., 2000). The other HTH motif in the Rob monomer contacts the DNA backbone.

The classic HTH motif consists of two helices, packed at angles of  $\sim 120^\circ$ , joined by a tight, four-residue turn in which a glycine is usually found in the second position. The HTH motif alone is insufficient for independent folding, so a third  $\alpha$ -helix stabilizes the motif. The second helix in the HTH motif is termed the "recognition helix", as it is inserted into the major groove of the DNA and is critical for specificity.

The HTH motif can be oriented in many ways in the major groove, even within homologous families. This variation in positioning adds to the variety and structural complexity of possible contacts with DNA. For example, the recognition helix often lies with its axis parallel to the edges of the bp to be recognized, as it does in  $\lambda$  cI (Figure 1.3A), where it can make many direct contacts to bases (Jordan and Pabo, 1988). In contrast, in TrpR, the N-terminal end of the recognition helix points into the major groove (Figure 1.3B), with its axis almost perpendicular to the DNA axis. The TrpR recognition helix makes no direct contacts to the bases within the cognate DNA site (Otwinowski et al., 1988). Instead, TrpR uses water-mediated contacts and indirect readout to specifically bind DNA. The precise position and orientation of the HTH recognition helix determines which protein functional groups will come close to the DNA bases.

A fabulous example of the diversity of the HTH motif is found in bacteriophage  $\lambda$  Cro and cI proteins. Both proteins are homodimers that use one HTH motif per monomer to recognize the same DNA sequences (Albright and Matthews, 1998a; Jordan and Pabo,

1988). It is remarkable that these proteins can differentiate between  $O_R1$  and  $O_R3$  and display opposing preferences for the two operators, but they are still capable of binding both operators. Cro and cI orient their HTH motifs in different ways with respect to the DNA (Figure 1.4, (Albright and Matthews, 1998b)). In Cro, the recognition helices of the two HTH motifs are 29 Å apart from center to center, tilted by 35° with respect to each other, and they are parallel to the major groove (Figure 1.4B). In contrast, the cI recognition helices are 34 Å apart, parallel to each other, and inclined to the major groove (Figure 1.4A). The recognition helices of each protein are sterically hindered from taking the same positions along the DNA. The symmetries of Cro-DNA and cI-DNA complexes are also dissimilar. Cro binds the 17 bp operator with two-fold symmetry, contacting the 7 outermost bp on each end. All direct base contacts are made from the Cro HTH motif and result in a 40° bend in the operator. Unlike Cro, cI—which also contacts 14 bp of the 17 bp operator—makes additional asymmetric contacts using the N-terminal arm from one monomer to interact only with the consensus half of the operator. The arm of the other monomer is largely disordered and does not make observable contacts to the nonconsensus half-site of the operator. Moreover, the DNA is only slightly bent in the cI-DNA structure.

In addition to variances in symmetry, Cro and cI have very different tertiary structures to which their HTH motifs are attached. While the cI dimer interface is made up of  $\alpha$  helices that do not appreciably change their relative positions upon binding DNA, the Cro dimer interface is forged from two antiparallel  $\beta$ -strands that can twist to allow flexibility. In fact, when Cro binds DNA, the subunits rotate 55° with respect to each other (Brennan et al., 1990).

The N-termini of the HTH recognition helices of both proteins begin with the sequence Gln-Ser, and these two residues make similar contacts with bp 2 and 4, two of the six bp that are invariant among  $O_R$  and  $O_L$  operator half-sites (Table 1.2). This is where the similarities end. Even within the three  $O_R$  sites there are significant differences. Bp 3' is the most critical for distinction between  $O_{R1}$  and  $O_{R3}$ . The two operators differ only at bp 3', 5' and 8' (Table 1.2). Whereas  $\lambda$  cI binds  $O_{R1}$  with higher affinity, Cro prefers  $O_{R3}$ . The middle residue of the recognition helix of each protein is responsible for interactions with bp 3'. In Cro, this residue is Asn31, while in cI it is Gly48. Cro prefers to bind DNA containing Thy(+3) (Takeda et al., 1989), as this allows the 5-methyl group of Thy to make a favorable van der Waals contact with the planar face of Asn31. Asn31 is also immobilized by hydrogen bonds with other protein residues and the phosphodiester backbone, thus placing it in position to make a favorable van der Waals interaction with the thymidine. TA is favored over AT at this position by the ability of Tyr26 to make van der Waals contacts to Thy(-3) if it is present. In cI, Gly48 is too far (4.2 Å) from Thy(+3) to make a strong van der Waals contact in the consensus half of the operator, and this correlates well with observations that cI makes no bp discrimination at this position (Sarai and Takeda, 1989). However, due to the asymmetry of the cI-DNA interaction, this interaction is even weaker (5.7 Å) in the nonconsensus half-site.

A subgroup of the HTH superfamily is the LacI/GalR family of transcription regulators. One member of this family, PurR, is the subject of the majority of experiments described in this thesis. PurR uses HTH motifs and "hinge helices" to make base-specific contacts in the major and minor grooves. Compared to most HTH proteins,

the HTH motif of PurR is reversed with respect to the dyad axis of the DNA. PurR also uses its hinge helices to bind the minor groove, in which two antiparallel helices (one from each monomer) come together only when PurR's small molecule corepressor has bound. These helices pry apart the minor groove, using "leucine levers" to widen the minor groove and induce a  $\sim 50^\circ$  kink in the DNA. The biochemistry and structures of PurR will be discussed in greater detail in a later section.

Other HTH motif subfamilies exist that will not be discussed in detail here, such as the homeodomain proteins. The first homeodomain structures solved were of AntP, engrailed, and Mat $\alpha$ 2 (reviewed in (Luisi, 1995)). These structures revealed that homeodomain proteins have substantially longer recognition helices than typical HTH domains, with more basic residues at their C-termini, resulting in an altered orientation of the helices on the DNA. Because homeodomains have been found exclusively in eukaryotes to date, they do not fit within the scope of this thesis.

### Winged helix proteins

Winged helix (WH) proteins are a large and diverse subgroup of the HTH family (reviewed in (Brennan, 1993)). The typical WH motif contains two wings, which are extended loop structures, in addition to three  $\alpha$ -helices and three  $\beta$ -strands in the following order: H1-S1-H2-H3-S2-W1-S3-W2 (Gajiwala and Burley, 2000). H2, H3, and the sequence between them make up the HTH motif, where H3 is the recognition helix, typically flanked by the wings, resembling a butterfly. The first structures of WH proteins were solved for the eukaryotic protein hepatocyte nuclear factor 3 (HNF-3) (Clark et al., 1993) and histone H5 (Ramakrishnan et al., 1993), but the motif appears to

be common across all kingdoms. Some of the prokaryotic members of the WH subfamily for which structures are available include CRP (mentioned previously), OmpR, a member of a two-component regulatory pathway involved in response to osmotic pressure (Martinez-Hackert and Stock, 1997), and BmrR, a multidrug binding member of the MerR family of transcription regulators (Heldwein and Brennan, 2001). Eukaryotic examples will be mentioned only when germane to the discussion.

In WH proteins, H3, the recognition helix of the HTH motif, lies in the major groove and makes most of the sequence-specific contacts with DNA. An exception to this rule is found in the human WH protein RFX1, in which the recognition helix lies over the minor groove, making a single DNA contact, while W1 makes extensive contacts with the major groove of DNA (Gajiwala et al., 2000). The length of the turn connecting the two helices of the HTH motif varies in WH proteins, allowing the crossing angles to vary from the  $\sim 120^\circ$  seen in canonical HTH proteins. For example, the biotin repressor, BirA, is a WH protein that has a crossing angle of  $\sim 100^\circ$  (Wilson et al., 1992), while heterodimeric cell-cycle transcription factor E2F-DP has a crossing angle of  $150^\circ$ .

In hindsight, it has been recognized that one of the canonical HTH proteins, CRP, is also a WH protein (Figure 1.5). CRP was not originally defined as a WH protein because its HTH motif showed no disparity from others, such as  $\lambda$  Cro and cI. Moreover, the CRP HTH helices cross each other at  $\sim 120^\circ$ , and only one contact is made between a distal wing residue and the phosphodiester backbone of DNA.

Several CRP-DNA complex structures have been solved to date (Parkinson et al., 1996; Schultz et al., 1991). The first, solved in the Steitz laboratory in 1991, was to 3.0

Å resolution. The second, solved by Helen Berman's group, was to 2.5 Å resolution and utilized a DNA oligonucleotide slightly altered from the first. There are differences between the two structures in specific protein-DNA contacts, the overall structure of the DNA, and in the regions that are visible, but the overall structures are the same. The CRP-bound DNA is bent 90°, the result of two very sharp kinks of ~40° each. These kinks are highly localized between Thy(-5)pGua(-6) and Thy(5')pGua(6') (Figure 1.5B), and pyrimidine-purine steps are known to kink most readily (Dickerson, 1998). Operator site recognition is effected by helix residues Arg180, Glu181, and Arg185 from each subunit, which make hydrogen bonds with DNA base edges. The wing in the CRP structure does not have a large role in recognizing DNA, but probably does help to position the HTH motif in the major groove. One wing residue, Lys26, does make two hydrogen bonds to Cyt(-9') phosphate groups. In a complex with straight DNA, these interactions might not be possible, so it may be that these contacts contribute to stabilizing the sharply bent DNA conformation of the complex structure.

Another WH protein, OmpR, and the osmosensor histidine protein kinase, EnvZ, allow cells to regulate differential expression of the major outer membrane porin proteins, OmpF and OmpC, in response to changes in osmolarity (Kato et al., 1989). At conditions of high osmolarity, phosphorylation of OmpR by EnvZ enhances its DNA binding activity so that transcription of *ompF* is repressed and that of *ompC* is activated. The N-terminal domain of OmpR contains the highly conserved aspartate residue that is phosphorylated (Delgado et al., 1993), while the C-terminal domain is responsible for binding cognate DNA and also likely interacts with αCTD of RNAP (Tsung et al., 1989). When the *E. coli* OmpR DNA binding domain structure was determined, OmpR was

shown to have a WH motif (Figure 1.6) (Martinez-Hackert and Stock, 1997). The OmpR C-terminal domain three helices superimpose well with BirA, yielding an rmsd of 1.05 Å.

Interestingly, the turn in the OmpR HTH unit is actually a 10-residue loop instead of a 4-residue turn. The W1 wing is present as the loop connecting  $\beta$ -ribbon strands B6 and B7. This hairpin is connected by a relatively long loop (8 amino acids, *c.f.* 2 in BirA) to the C-terminal end of the recognition helix. OmpR also has an N-terminal 4-stranded  $\beta$ -sheet not found in other WH protein families (Figure 1.6). This sheet is also found in the related protein PhoB (Okamura et al., 2000), and it appears to take the place of W2. Therefore, both the wings and the turn in the HTH motif differ in OmpR from those of other WH proteins. The OmpR-DNA complex structure has not yet been solved.

Due to the high degree of conservation of amino acid sequence identity, it has been proposed that all DNA binding members of two-component response regulatory systems will have similar overall folds to that of OmpR. The recent NMR structure of PhoB is very similar to OmpR, supporting this hypothesis (Okamura et al., 2000). The structure of NarL, a two-component regulatory partner protein homologous to OmpR, has been solved in its full-length, nonphosphorylated form (Baikalov et al., 1996). In this structure, the DNA binding domain is covered by the regulatory domain, precluding DNA binding, suggesting a possible regulatory mechanism for this family of proteins.

The recent structure of the MuR-DNA complex has revealed yet another role for the wing: binding in the minor groove (Wojciak et al., 2001). MuR promotes lysogeny of the Mu bacteriophage by binding the Mu transposon's enhancer element, IAS (for Internal Activation Sequence), thus repressing genes required for lytic growth and blocking the phage encoded transposase MuA from binding the IAS. Both MuR and



MuA IAS binding domains have been predicted to have WH motifs. The structure of the MuA DNA binding domain in the absence of DNA was solved by NMR and showed that MuA is a WH protein (Clubb et al., 1994). Addition of DNA resulted in chemical shifts in the recognition helix, the wing, and the beginning of the  $\beta$ -strand C-terminal to the wing. The structure of the N-terminal MuR DNA binding domain, residues 1-84, in complex with DNA, was recently solved by NMR (Wojciak et al., 2001). MuR has an altered topology from canonical WH proteins: B1-H1-T-H2-B2-W-B3-H3, where H1 and H2, separated by a seven-residue turn, form the HTH motif (Figure 1.6D). DNA sequence specificity is largely achieved through contacts in the major groove from H2, the recognition helix. This helix is oriented roughly perpendicular to the DNA helical axis. Base-specific interactions are made by residues at the N-terminus of the recognition helix. The novel aspect of the MuR DNA binding domain is that, upon binding cognate DNA, the wing undergoes a disorder-to-order transition and becomes immobilized in the minor groove, where it makes direct hydrogen bonds to nucleotides. Lys53 lies on the floor of the minor groove; its terminal amine donates hydrogen bonds to the invariant O2 and N3 atoms of Thy25 and Ade12. Wing residue Lys56 also forms hydrogen bonds in the minor groove, to the O2 and O4' oxygens of Thy28 and Gua29, respectively. Other contacts are made, but differentiating between bases in the minor groove is not as inherently specific a process as differentiating in the major groove (Figure 1.2). Therefore, the likely role of the wing of MuR is to help position the protein on DNA.

Another recently reported WH structure is that of the *B. subtilis* BmrR-TPP-DNA complex (Figure 1.7A, (Heldwein and Brennan, 2001). BmrR is a MerR family member that activates the transcription of a gene encoding a multiple drug transporter gene upon

binding lipophilic cations, thus leading to multidrug resistance. Members of the MerR family of transcription regulators can function as activators or repressors depending on whether coactivating ligands are bound (reviewed in (Summers, 1992)). Like all MerR family proteins, BmrR binds as a homodimer to a promoter that has a 19 bp spacer between its -35 and -10 elements, 2 bp longer than usual, nonrestrictive promoters. This longer spacer places these elements on different faces of the DNA helix,  $\sim 70^\circ$  out of phase (Figure 1.7B). The crystal structure of BmrR revealed that each monomer has 3 domains: an N-terminal DNA binding domain, a long helical linker, and a drug binding domain. The long linker helices form an antiparallel coiled coil in the homodimeric structure, and the drug and DNA binding domains make up part of the dimerization interface (Figure 1.7A).

The DNA binding domain has the topology: B1-H1-H2-B2-B3-H3-H4, where B represents  $\beta$ -strands and H represents helices. The DNA binding domain is forged from a HTH motif (H1-H2) and two wings (W1 and W2). The recognition helices contact the major groove, with each helical axis roughly perpendicular to the local DNA helical axis. Helix H1 makes contacts to two DNA backbone sugars, and the recognition helix, H2, makes the remainder of the HTH contacts. W1 is made from the B2-B3 hairpin and the loop connecting them. This first wing makes multiple positioning contacts to DNA backbone phosphate and sugar groups. W2 is actually another HTH motif (H3-turn-H4). The helices are less crossed than in most HTH motifs, but this is common in the WH family, as mentioned previously for BirA (Wilson et al., 1992).

The structure of the DNA is highly distorted in the BmrR-drug complex. The promoter is bent away from the protein, toward the major groove, by  $\sim 50^\circ$ . The AT bp at

the pseudo dyad of the promoter breaks its hydrogen bonds such that the adenine and thymidine slide away from each other. In essence, the DNA "bunches up" in the middle, effectively shortening the promoter by  $\sim 5$  Å. This DNA distortion brings the -35 and -10 elements into phase on the DNA (Figure 1.7B), thus making both promoter elements accessible to RNAP. Surprisingly, in this drug-bound structure, there are no base-specifying contacts between BmrR and its cognate DNA. Either DNA binding specificity must be achieved by indirect readout, or only apo BmrR must make specific contacts, or the contacts must not be seen in this structure due to the statistical disorder of the DNA. The BmrR-drug-DNA structure will be discussed in more detail later in this chapter and in Chapter 6, as it relates to my studies with the *P. mirabilis* MerR family member, PMTR.

### The $\beta$ -ribbon motif

The use of  $\beta$ -strands to bind DNA is less common than use of  $\alpha$ -helices. There are two known classes of prokaryotic DNA binding proteins that bind their cognate DNA using  $\beta$ -sheet motifs. The first class includes the MetJ, Arc, Mnt, and CopG repressors, which all use an antiparallel  $\beta$ -ribbon for DNA binding. The cocrystal structures of *E. coli* MetJ, Arc, and CopG repressors in complex with DNA have been solved (Gomis-Rüth et al., 1998; Raumann et al., 1994; Somers and Phillips, 1992). The structure of Mnt has been solved in the absence of DNA (Burgering et al., 1994). The second class of  $\beta$ -sheet DNA binding motif consists of the IHF, HU, and TF1 proteins. These proteins use a  $\beta$ -sheet and a pair of  $\beta$ -arms to recognize DNA. The structure of an *E. coli* IHF-DNA complex has also been solved (Rice et al., 1996). The structure of *B.*

*stearothermophilus* HU protein was solved in the absence of DNA (Tanaka et al., 1984; White et al., 1999), and that of phage SPO1 TF1 has also been reported (Jia et al., 1996). MetJ, Arc, and IHF will be discussed briefly. There is at least one other class of  $\beta$ -sheet motif used to bind DNA in eukaryotes. The TATA-binding protein (TBP) employs a 10-stranded  $\beta$ -sheet in the shape of a saddle to bind DNA in the minor groove (Kim et al., 1993a; Kim et al., 1993b).

The *E. coli* methionine repressor, MetJ, regulates genes in the methionine biosynthetic pathway by binding at least eight different *met* operators. MetJ is activated to bind DNA when it has noncooperatively bound two S-adenosyl methionine (SAM) molecules in a mechanism of feedback inhibition, as SAM is generated from methionine. The "met box" consensus sequence is AGACGTCT (Wu et al., 1993). These boxes are always found in tandem, with two boxes minimally required for efficient operator control (*ibid.*). MetJ-DNA binding is highly cooperative. The MetJ-DNA cocrystal structure showed that two identical subunits form an intrinsically intertwined homodimer with two-fold symmetry (Figure 1.8A, (Somers and Phillips, 1992). Each subunit is made up of one  $\beta$ -strand, three  $\alpha$ -helices (A, B, and C), and long connecting loops between the helices. The  $\beta$ -strands come together to form a two-stranded antiparallel  $\beta$ -sheet, or " $\beta$ -ribbon". This  $\beta$ -ribbon motif binds in the major groove of DNA. The two symmetry-related B helices form the hydrophobic core and also bind the two molecules of SAM. The core structure does not change appreciably upon binding either SAM or DNA. The structure also revealed two MetJ dimers bound to two adjacent met boxes of the 19 bp oligonucleotide used for cocrystallization. Dimer-dimer interactions are mediated by the "A" helices, which lie above the major groove.

The met box DNA is B-form but bent  $\sim 25^\circ$  toward the major groove. The resulting compression of the major groove allows the close packing of the  $\beta$ -ribbon within the groove. Solvent and ion molecules are excluded from the interface, and MetJ makes numerous hydrogen bond and salt-bridge interactions. Side chains of the  $\beta$ -strands provide two direct hydrogen bonds (Thr25 O $\gamma$ 1-Ade3 N7 and Lys23 N $\zeta$ -Gua10' O6) and one water-mediated interaction (also from Lys23) per monomer to the bases. These constitute the only base-specific hydrogen bonds observed in the complex. Ade3 cannot be replaced with Thy, as its 5-methyl group would clash sterically with Thr25. Cyt4 may make water-mediated contacts not observed in the 2.8 Å structure. The central Thy8-Ade9 step, comprising the end of one met box and the beginning of another, is heavily overwound, and an adjacent phosphate is contacted by the N-terminus of the B helix. TpA steps have weak base stacking energies because they have small bp overlap (Dickerson, 1998), accounting for the indirect preference for these bases at this position.

The bacteriophage Arc repressor also uses an antiparallel  $\beta$ -ribbon to bind DNA, with a few key distinctions from the MetJ structure (Figure 1.8B, (Raumann et al., 1994). The Arc repressor is involved in the switch between lysis and lysogeny in the *Salmonella* bacteriophage P22 by controlling the transcription of the *ant* gene. Arc binds DNA in a manner similar to that of MetJ, however Arc-DNA interactions are far less symmetrical. The *ant* promoter contains two consensus TAGA boxes with five bp between them, related by two-fold symmetry, that are recognized differentially by each monomer. Also, the core domain of Arc undergoes conformational changes in the  $\beta$ -sheet recognition motif, while no change in conformation of MetJ is observed upon binding DNA. Arc does form a cooperative dimer of dimers on DNA, as MetJ does, but uses the loop

between helices A and B to make these interactions, rather than the entire length of helix A that is used by MetJ. Arc also distorts the structure of the DNA more than MetJ, bending it by  $\sim 50^\circ$ . The average twist of the DNA is  $32.3^\circ/\text{bp}$ , which yields 11.1 bp/turn. Near the  $\beta$ -ribbons, the major groove is widened to 14.3 Å. At the center of the operator, the minor groove is only 1.2 Å wide.  $\alpha\text{-NH}_3^+$  groups from each subunit of the protein minimize the repulsion of the two phosphates that become closest across the minor groove.

Another class of prokaryotic DNA binding protein that uses  $\beta$ -strands to bind DNA has three members whose structures have been solved: *E. coli* IHF, *B. stearothermophilus* HU, and bacteriophage SPO1 TF1. The crystal structure of IHF in complex with 35 bp of DNA has been solved (Rice et al., 1996). The structure of apo HU has also been determined (Tanaka et al., 1984; White et al., 1999). HU is an architectural DNA binding protein that binds DNA nonspecifically to introduce negative supercoiling within the circular DNA molecule, and its overall structure closely resembles that of IHF. The rmsd between IHF and HU core residue  $\text{C}\alpha$  atoms is 1.0 Å. The NMR structure of SPO1 TF1 reveals that it is similar to both IHF and HU (Jia et al., 1996).

The integration host factor, IHF, was discovered as a host factor for bacteriophage  $\lambda$  integration, but several other roles for IHF in processes requiring higher order protein-DNA complexes have been delineated. Its main function with respect to transcriptional regulation is activation by binding upstream of many  $\sigma^{54}$ -dependent promoters (Dworkin et al., 1997). However, the role of IHF is always architectural in that it introduces a sharp bend in the DNA (Rice et al., 1996), which facilitates the interaction of other components

in these higher order protein-DNA complexes. IHF is an  $\alpha/\beta$  heterodimer composed of two 10 kDa subunits, which are 30% identical in primary sequence. The DNA binding site is not well conserved, but certain sequences are preferred over random DNA. IHF binds DNA using long  $\beta$ -arms, which interact exclusively with the minor groove (Figure 1.9). This leaves the major groove open for interaction with other proteins, a feature probably key for IHF function. The DNA is bent 160-180°, such that it wraps around the small protein. The bending is primarily localized at two kinks, nine bp apart. A proline from the tip of each  $\beta$  arm is intercalated between the Ade37pAde38 step at these kinks. This proline is conserved in all known members of the IHF/HU family. Prolines may stabilize the turn at the tip of the  $\beta$ -arms, but they may also make favorable hydrophobic contacts with the DNA bases. To bend DNA so extremely, IHF must somehow overcome the forces that normally favor straight DNA, i.e. symmetric repulsion of phosphate backbone charges and favorable base stacking energies. IHF positions positive charges on the inside of the bend, to neutralize phosphate backbone repulsion, and it inserts a hydrophobic residue between bp at each large kink to undo the energetic destabilization of disrupting base stacking.

IHF prefers to bind DNA with an A-tract in the left half-site (Yang and Nash, 1989). A-tracts are often straight DNA with a narrow minor groove and high propeller twist. The IHF-DNA structure has one of these A-tracts and implies that this is likely another example of indirect readout. In fact, this region of DNA, when superimposed with the structure of an A-tract dodecamer (Nelson et al., 1987), has an rmsd of 0.6 Å for all atoms. The narrow minor groove fits well with the tripartite protein clamp that is forged from the N-termini of two  $\alpha$ -helices and a turn between two  $\beta$ -strands. Indirect

readout, recognition of DNA by its sequence-dependent conformation or deformability, is likely to be particularly important to minor groove binding proteins because of the relative lack of distinguishing features in the minor groove.



## The *E. coli* purine repressor, PurR

### The LacI/GalR family

The *E. coli* lactose repressor protein, LacI, is the prototype for genetic regulatory proteins. Its existence was predicted by Jacob and Monod in 1961 (Jacob and Monod, 1961), however, the protein itself was not identified for another five years (Gilbert et al., 1966). The allosteric model for LacI-mediated gene regulatory control was soon borne out by studies with the *lac* operon (Monod et al., 1965). In this model, LacI binding to specific operator DNA sequences was modulated by the abundance of particular sugars. LacI regulates the transcription of genes responsible for lactose metabolism (reviewed in (Matthews and Nichols, 1998)): *lacZ* encodes  $\beta$ -galactosidase, *lacY* codes for *lac* permease, and *lacA* is the gene for a transacetylase. When lactose is available, *lac* permease transports it in and  $\beta$ -galactosidase cleaves as well as subsequently converts the sugar to the natural inducer,  $\beta$ -1,6-allolactose. LacI binds this sugar, which causes a conformational change resulting in decreased affinity of LacI for specific DNA. When lactose levels fall, so do allolactose levels, ultimately causing the inducer to dissociate from LacI. LacI undergoes a conformational change and again binds operator DNA with high affinity. LacI has been extensively characterized genetically, physically, biochemically, and recently by X-ray crystallography. In one paper, over 4000 single amino acid replacements were analyzed phenotypically (Markiewicz et al., 1994)! The resulting phenotypes were classified as  $I^S$ , unable to respond to inducer but capable of binding DNA,  $I^-$ , no longer able to repress transcription of *lac* genes, or  $I^+$ , in which repression increased with increasing concentrations of inducer.

Since the discovery of LacI, at least 24 other LacI/GalR family members have been described (Nguyen and Saier, 1995). Each of these proteins has a highly conserved DNA binding domain (DBD), residues ~1-60, and a less conserved effector binding domain (CBD, named for the corepressor binding domain of PurR), residues ~61-341. The DBDs are all predicted to have HTH motifs, and the C-terminal domains show high structural homology to the bacterial periplasmic binding proteins, PBP (Schumacher et al., 1993). The operator DNA sequences recognized by LacI/GalR family members are usually 16-18 bp and show limited conservation. In most of these proteins, binding of a ligand to the larger C-terminal CBD causes a conformational change, which causes protein dissociation from the DNA. Exceptions are found in the *E. coli* purine repressor, PurR (Rolfes and Zalkin, 1990b), and the catabolite control protein, CcpA (Jones et al., 1997). PurR must first bind its corepressor ligand before it can bind specific DNA with high affinity. As will be discussed later, this may be because PurR controls the expression of genes involved in anabolic pathways, while other LacI/GalR family members are induced by molecules related to the catabolic pathways they control. A phylogenetic tree constructed by Nguyen and Saier (Figure 4.1A) shows that almost all of the evolutionary branches are equal in length and radiate from a single point (Nguyen and Saier, 1995). This suggests that not long (evolutionarily) after the DNA binding domains and ligand binding domains were fused, gene duplication events were selected for that gave rise to the current, functionally dissimilar members of the LacI/GalR family, after which time few additional gene duplications occurred. Interestingly, PurR and RbsR, the ribose repressor, evidently diverged from each other at some later evolutionary time. This fact will be revisited later in this section and again in Chapter 5 of this thesis.

### Regulation of purine biosynthesis by PurR

In *E. coli*, purine nucleotides can be derived by salvage pathways from exogenous purine pools, or they can be synthesized *de novo* (reviewed in (Zalkin and Nygaard, 1996)). *De novo* synthesis of IMP, which is readily converted to AMP or GMP, is accomplished in eleven steps by enzymes scattered about the *E. coli* genome in nine separate loci that make up the *pur* regulon (Figure 1.10A). Operator DNA sequences bound by PurR are shown in Figure 1.10B. Beginning with PRPP (5-phospho- $\alpha$ -D-ribosyl-1-pyrophosphate), the first committed step in IMP biosynthesis is catalyzed by glutamine PRPP amidotransferase, encoded by the *purF* gene. In response to binding of the physiological corepressors, hypoxanthine or guanine, PurR represses the transcription of this gene by ~17-fold *in vivo* by binding the 16 bp *purF* operator, which is situated over the -35 element of the promoter (Figure 1.10C). The Gln PRPP amidotransferase is also subject to feedback inhibition by purine nucleotides, with GMP inhibiting at lower concentrations than AMP. The other genes involved in IMP biosynthesis are also downregulated by greater than 10-fold by PurR binding to similar operator sequences. Among these promoters (Figure 1.10B), PurR binding to the *purF* operator is the best characterized.

PurR not only regulates genes involved in IMP synthesis, but also the expression of genes encoding the enzymes that convert IMP to GMP (*guaAB*) and IMP to AMP (*purA* and *purB*), although these genes are regulated only 2-5 fold by PurR (Meng et al., 1990). Differences in regulation can be attributed to differential operator affinity and the position of the operator with respect to the promoter (Figure 1.10C). PurR binds an

operator within the coding region of *purB*, so it is expected to block transcription elongation in this gene (He et al., 1992). The lower repression of genes responsible for conversion of IMP to GMP and AMP reflects cells' requirement to utilize salvaged purines when *de novo* IMP synthesis is fully repressed. The expression of PurR is autoregulated 2-3 fold (Rolfes and Zalkin, 1990a). There are two *purR* operator sites, O<sub>1</sub> and O<sub>2</sub>, that are located 103 and 191 bp downstream of the *purR* transcription initiation site. PurR binds independently to the O<sub>1</sub> operator with higher affinity than for O<sub>2</sub>. Both operators contribute to the total 2-3 fold autoregulation.

PurR also is involved in the purine-mediated regulation of other genes that are related to nucleotide metabolism (Choi and Zalkin, 1990; He et al., 1993; Steiert et al., 1990). PurR regulates *pyrC* and *pyrD*, which are involved in the *de novo* synthesis of pyrimidine nucleotides, *codBA*, involved in cytosine transport and salvage, and *prs*, which encodes PRPP synthetase. PRPP synthetase catalyzes the conversion of ribose-5-phosphate to PRPP. PRPP is required for the synthesis of purine, pyrimidine, and nicotinamide nucleotides, and the amino acids histidine and tryptophan. The *prs* gene is regulated by other transcription factors in addition to PurR. Additionally, PurR regulates *glyA* and the *gcv* operon, involved in glycine synthesis and catabolism and the synthesis of one carbon units. PurR regulates *speA*, involved in polyamine synthesis, and *glnB*, required for regulation of the *gln* operon and thus glutamine synthesis.

### Structures of PurR

PurR is a 341 amino acid residue, 38 kDa protein that is always found as a homodimer. The crystal structures of the holo complex of PurR-purine-DNA and the apo

PurR "core" –the corepressor binding domain alone–have been solved (Schumacher et al., 1995; Schumacher et al., 1994a; Schumacher et al., 1997). The N-terminal DNA binding domain (DBD, residues 1-60), consists of a HTH motif (residues 4-23), followed by a short loop, another helix, and another loop, making up a globular subdomain (Figure 1.11A). This is followed by a second DNA binding motif, the "hinge" helix (residues 48-56), which is followed by four extended residues that link the N- and C-terminal domains. The C-terminal corepressor binding domain (CBD) can be further divided into two structurally similar subdomains, termed the "N-subdomain" and the "C-subdomain". The N-subdomain has a 6-stranded parallel  $\beta$ -sheet surrounded by 4  $\alpha$ -helices, while the C-subdomain has a 5-stranded parallel  $\beta$ -sheet enclosed within 4  $\alpha$ -helices and two single turns of  $3_{10}$  helices. Three strands cross over to connect the CBD subdomains. The corepressor binding pocket is situated between the N- and C-subdomains of the CBD, involving residues from both subdomains and from the crossover strands.

The dimeric CBD shows striking structural similarity to the monomeric periplasmic binding proteins (PBP) (Quioco and Ledvina, 1996), particularly the ribose binding protein, RBP (Figure 1.12) (Mowbray and Cole, 1992). The root mean-squared deviation (rmsd) for C $\alpha$ s between PurR and RBP is 1.8 Å for the N-subdomains and 2.3 Å for the C-subdomains. Although the PurR structures were the first solved for any LacI/GalR family member, subsequent structures have been reported for LacI in induced and repressed states (Bell and Lewis, 2000; Friedman et al., 1995; Lewis et al., 1996) and the effector binding domain of TreR, the *E. coli* trehalose repressor, bound to inducer and noninducer (Hars et al., 1998). These structures all take very similar topologies (Figure 1.12), with the exception that LacI has a C-terminal leucine zipper-like tetramerization

domain, which folds into a four-helix bundle. In addition to RBP, at least 12 other PBP structures have been solved, including the arabinose binding protein (Quiocho and Vyas, 1984) and leucine/isoleucine/valine binding protein (Sack et al., 1989). Solution structures have also been reported for the DNA binding domains of PurR (Nagadoi et al., 1995), LacI (Slijper et al., 1996; Spronk et al., 1999), and FruR (Penin et al., 1997).

### DNA recognition

DNA binding by PurR is effected by two DNA-binding elements. The first is the HTH motif, which is similar to those of many other DNA binding proteins, but has a few of its own idiosyncrasies. The HTH motif in PurR replaces the "invariant" glycine in the second position of the turn with an asparagine, but this has very little effect on the structure; the rmsd between the HTH of PurR and that of phage  $\lambda$  cI is 0.55 Å. In PurR, the N-terminus of the recognition helix points into the major groove and makes extensive contacts. There are 13 observed protein-DNA backbone interactions, and five residues contribute direct or water-mediated base contacts in the major groove. These base contacts account for much of the affinity for the PurR consensus DNA sequence. One key specificity-determining contact is made from Arg26, located just after the HTH motif, to Gua4, which is conserved in all 42 known cognate DNA half-sites bound by PurR. The orientation of the PurR HTH elements is reversed with respect to the dyad axis of the DNA, compared to most HTH proteins. This reversed orientation is found in the LacI-DNA and FruR-DNA structures (Lewis et al., 1996; Penin et al., 1997) and is likely to be common across the family. The tetracycline repressor HTH also binds its

cognate DNA with a reversed orientation, although it is not a LacI/GalR family member (Orth et al., 2000).

The second DNA binding elements are the dyad-related hinge helices. These hinge helices are disordered in the absence of corepressor and cognate DNA, as was shown by proteolysis and NMR studies (Choi and Zalkin, 1994; Nagadoi et al., 1995). In the PurR-hypoxanthine-*purF* operator complex, the hinge helices are juxtaposed and engage in van der Waals contacts between both Val50 $\leftrightarrow$ Val50' and Leu54 $\leftrightarrow$ Leu54'. These helices bind in the minor groove at the central CpG step of the operator (bp 9-9'), expanding the minor groove from 5.7 Å (B-form DNA) to 10.3 Å. This is accomplished by the insertion of Leu54 and Leu54' –the "leucine levers"– into the CpG step, resulting in a large roll angle of 45°, kinking of the DNA by ~50°, and the unstacking of the base pairs of this step. This DNA deformation is highly localized, with bp 3-7 having near B-form DNA parameters (data not shown). The structure of LacI reveals that it, too, uses leucine levers in an analogous fashion (Lewis et al., 1996; Slijper et al., 1996). This leucine is conserved across the LacI/GalR family, with three exceptions: CytR, which has a valine at this position, and TreR and CscR, which have methionines. An L54V mutant PurR showed no *in vivo* activity, nor did substitutions with K, S, W, T, or R (Arvidson et al., 1998). PurR L54M was able to regulate *purF-lacZ* expression *in vivo*, and it bound *purF* operator with only 2-fold reduced affinity *in vitro*. The crystal structure of the L54M-hypoxanthine-*purF* operator complex was solved and revealed that the Met54 side chain, while taking strained  $\chi_1$  and  $\chi_2$  torsion angles, was able to penetrate deeper into the minor groove but did not span the width of the groove as completely as the branched leucine. In summary, DNA recognition in PurR, and likely throughout the LacI/GalR

family, is achieved by a combination of base-specific contacts and indirect readout (sequence dependent DNA deformability).

### Allosteric mechanism

When we compare the PurR holorepressor and aporepressor structures, we see that the C-subdomains of the CBD do not move with respect to each other (Figure 1.11). The C-subdomains of the corepressor-bound and corepressor-free structures can be overlaid with an rmsd of less than 0.4 Å. However, in the apo PurR-CBD structure, the N-subdomains undergo hinge-bending rotations of 17° and 23° relative to their positions in the corepressor-bound form, with rmsds of 1.7 and 1.6 Å between holo and apo forms of each monomer. This rotation alters the dimer interface and disengages the hinge helices from the minor groove. Although the CBD structure does not include the hinge helices (residues 53-59 and 188-193 are disordered), the Lys60 Cα - Lys60' Cα distance increases from 17.3 Å in the holorepressor to 20.8 Å in the aporepressor. Rearrangement of the corepressor binding pocket upon binding is the initial step in transduction of the intramolecular signal that corepressor has bound from the CBD to the DBD. Close-up views of the corepressor binding pocket, situated between the N- and C-subdomains of the CBD, are shown in Figure 1.13, where A is the holo (hypoxanthine-bound) structure and B is the apo (corepressor-free) structure.

A key residue in this allosteric "switch" mechanism of PurR is Trp147. In the open, corepressor-free form of PurR, Trp147 is found in the corepressor binding pocket, stacked against Tyr73 and Phe74, where it makes one hydrogen bond from its indole N9 to the Tyr73 hydroxyl group. Thus, Trp147 acts as a structural surrogate for corepressor.



However, Trp147 is not a functional surrogate for corepressor. Instead it appears to act as an intrasteric autoregulatory residue (Kobe and Kemp, 1999). Its contacts do not duplicate those made by hypoxanthine or guanine, yet it can be regarded as a "pseudo" ligand. Upon binding of a corepressor, Trp147 is displaced from the binding pocket. As part of the allosteric "switch", Trp147 rotates away from the pocket, pulling residues 148-159 with it, such that its indole side chain stacks instead against Tyr126. In Chapter 3 of this thesis, I report my work on PurR mutants W147F, A, and R, which involves further dissection of the key role played by this residue in intramolecular signalling. Indeed, mutation to Phe147 causes *in vivo* repression to weaken, whilst mutation to Ala147 and Arg147 causes a dramatic increase over wild-type repression levels. *In vitro* biochemical and structural analyses of these mutants are discussed in detail in Chapter 3.

#### Corepressor binding and specificity

The corepressor binding pocket is lined by residues making hydrogen bonds to the purine base that confer specificity due to their acceptor-donor restrictions. These hydrogen bonds, together with planar protein-style stacking interactions, allow PurR to bind its physiological corepressors with high affinity ( $K_{d(\text{gua})} = 2 \mu\text{M}$ ,  $K_{d(\text{hyp})} = 9 \mu\text{M}$ , (Choi and Zalkin, 1992). The Asp275 side chain carbonyl makes hydrogen bonds to the purine N9, the Thr192 side chain hydroxyl group makes a hydrogen bond to N7, and Arg190 makes two hydrogen bonds to the exocyclic O6 of guanine or hypoxanthine (Schumacher et al., 1994a; Schumacher et al., 1997). The Arg190 interaction has been shown to allow PurR to selectively bind 6-oxopurines by requiring that the exocyclic group at position 6 be a hydrogen bond acceptor (Lu et al., 1998b). This precludes binding of a 6-

aminopurine, such as adenine. However, by substituting Arg190 with Ala or Gln, corepressor specificity was expanded to include adenine and 6-methylpurine. A water-mediated interaction between the Glu222 carboxylate group, the Phe221 backbone carbonyl, and the exocyclic N2 of guanine has also been postulated to confer affinity for guanine over hypoxanthine and discrimination against xanthine, which has an exocyclic O2 (Schumacher et al., 1997). The Arg190 work will be reviewed in greater depth in Chapters 4 and 5 as it relates to work presented there. Work I have done on Glu222 mutants in collaboration with Drs. Dennis Arvidson, Maria Schumacher, and Fu Lu of the Brennan and Zalkin laboratories is presented in detail in Chapter 4. In that study we expanded corepressor specificity to include xanthine by site-directed mutagenesis of Glu222 to Ala, Gln, and Lys.

Finally, we wished to "evolve" the ligand binding pocket further. The residues that contact hypoxanthine or guanine in PurR are conserved with RbsR, with two exceptions. Conserved binding pocket residues include Tyr73, Phe74, Phe221, Glu222, and Asp275. The two nonconserved residues are Arg190 and Thr192, which are Lys and Pro in RbsR, respectively. The periplasmic ribose binding protein RBP also uses similar residues to bind ribose (Figure 5.2). Based on PurR similarities with RbsR and RBP, attempts to further manipulate corepressor specificity are presented in Chapter 5. We wanted to know how many residues had to be changed in order to dramatically alter ligand specificity. In those experiments, Arg190Lys and Thr192Pro mutations were designed to convert specificity more drastically from purines to sugars, with marginal success.

In summary, the *E. coli* purine repressor is a well-characterized system amenable to biochemical and structural studies. These experiments are relevant to our understanding of the function of transcription regulators. Allosterism, or intramolecular signal transduction, is common among DNA-binding proteins, however the mechanism by which it proceeds is not well understood at the atomic level for most proteins. In order to delineate the role of one residue that is key for the allosteric transition in PurR, Trp147, I have determined the binding affinities and structures of three mutant PurR proteins with substitutions at this position. Understanding how protein ligand binding sites achieve specificity is one of the most fundamental questions in structural biology, and testing our understanding is an essential part of the scientific method. By attempting to expand or evolve the specificity of the corepressor binding site in PurR, as was done with the E222, R190, and T192 replacements, we have tested our hypotheses regarding the contribution of individual residues to ligand binding.

### **Prokaryotic heavy metal-responsive transcription regulators**

Many heavy metals, such as Cu, Zn, Fe, Co, and Ni, are essential trace elements for bacteria (Ralston and O'Halloran, 1990). But these "trace nutrients" are toxic at micro- or millimolar concentrations. Nonessential heavy metals, such as Cd or Hg, are often toxic at even lower levels. Bacterial heavy metal uptake systems are not specific; they are designed to let in a broad range of ions at relatively high rates. For example, in *S. aureus*, Cd(II) enters the cell through the Mn(II) uptake system, while Zn(II) is imported via the Mg(II) system (Nies, 1992).

Prokaryotes have evolved two major systems of heavy metal resistance: active efflux of the toxic ion and enzymatic detoxification. Other mechanisms of resistance are possible, such as exclusion, sequestration, and reduced sensitivity of cellular targets (Bruins et al., 2000), but these are not often observed and have not been well characterized to date. For purposes of this thesis, "metalloregulatory" proteins shall be defined as those proteins that respond to changes in intracellular levels of specific metal ions by altering the levels of expression of genes that constitute a resistance system to that particular metal ion. All such proteins either bind both the metal ion and DNA directly or are the DNA binding member of a two-component regulatory system. I shall focus on the metalloreulatory proteins that respond to changes in specific intracellular metal concentrations and regulate expression of efflux and detoxification systems. These transcription factors may be put into four families based on sequence homologies and known functions: i) The MerR family; ii) the ArsR family, including CadC, SmtB, DtxR, and IdeR; iii) the two-component regulatory systems, for example CopRS and CzcRS (with similarity to OmpR/EnvZ and PhoBR); iv) other. Many proteins must be placed in

the "other" category until they are better characterized or until more proteins with similar modes of actions are identified. At present, our understanding of the regulation of most heavy metal resistance systems is muddled.

#### Metalloregulatory proteins that are not MerR family members

As mentioned previously, other groups of metalloregulatory proteins have been defined that have no homology to the MerR family. Several excellent reviews have compiled most of the known information regarding these systems (O'Halloran, 1993; Rosen, 1999; Silver, 1996).

The ArsR protein is one of two metalloregulatory proteins produced by the *ars* operon (*arsRDABC*), the other being ArsD. Both are repressors that bind to the same operator sequence in the *ars* promoter, forming a homeostatic regulatory control circuit that functions to maintain expression of genes conferring resistance to arsenite [As(III)] and antimonite [Sb(III)] (Chen and Rosen, 1997). ArsR is a high affinity DNA binding protein that responds to low concentrations of arsenical or antimonial compounds ( $< 10 \mu\text{M}$ ) by binding them and dissociating from its cognate DNA, thus relieving repression. This allows transcription of the *ars* operon as a polycistronic transcript. As levels of ArsD, the lower affinity repressor, increase, it binds to operator and transcription is again repressed. As As(III) and Sb(III) levels increase further, ArsD binds the metals and dissociates from the DNA. Little else is known about these two metalloregulatory proteins, save their homology to CadC, SmtB, and DtxR, and therefore that they are likely to be winged helix proteins.

CadC regulates the CadA resistance operon of *S. aureus* plasmid pI258 (Nies, 1992). CadA is a P-type ATPase that has been shown to transport cadmium into everted vesicles. CadC is a metalloregulatory protein and part of the CadCA operon. This was a disputed claim for some time, as CadC was initially hypothesized to have a "metal gathering" role by some groups. The role of CadC as a DNA binding protein has now been proven by the Silver group (Endo and Silver, 1995). Cd(II), Zn(II), Bi(III), and Pb(II) have all been shown to induce the release of CadC from cognate DNA. The footprint of CadC covers an inverted repeat region from -7 to +14 relative to the transcription start site of the *cadCA* operon. Other systems of cadmium resistance are used in other organisms (the *czc* system, for example (Nies and Silver, 1989).

The IdeR, SmtB, and DtxR proteins are members of the ArsR family whose crystal structures have been solved (Cook et al., 1998; Pohl et al., 1999; White et al., 1998). SmtB is a metalloregulatory protein that controls the expression of the *smtA* and *smtB* genes of the cyanobacterium *Synechococcus*. SmtA was the first described bacterial metallothionein, which acts by sequestering Cd(II), Zn(II), and Cu(II). SmtB represses transcription by binding *smt* operator/promoter DNA in the absence of Zn(II) and dissociating from DNA when Zn(II) is bound. The crystal structure of SmtB reveals that it has a winged helix DNA binding motif containing a standard HTH motif (Figure 1.14), (Cook et al., 1998). This motif overlays well with those of DtxR, HNF-3, and CRP. Modelling B-DNA onto the apo SmtB structure indicated that binding to both DNA binding motifs of the homodimer would likely require a bend of about 30° in the DNA helix. In derivatizing SmtB crystals with mercuric acetate for structure determination, four Hg(II) sites of low occupancy were found that might correspond to

the four postulated Zn(II) binding sites, although this is highly speculative. Two of the sites, formed by Cys61, Asp64, and His97 in each monomer, suggest a way by which Zn(II) binding could disrupt DNA binding, as they are located near the beginning of the HTH motif.

Another well characterized family of metalloregulatory proteins, which induce resistance operons in response to toxic levels of heavy metals, are the two-component regulatory system (or two-component signal transduction) proteins (Grosse et al., 1999; Munson et al., 2000). These proteins are similar to OmpR, discussed in the earlier winged helix DNA binding motif section. These regulators have a histidine kinase sensor partner that phosphorylates a conserved aspartate residue in the DNA binding protein in response to increased levels of heavy metal ions, much as EnvZ phosphorylates OmpR in response to changes in osmotic pressure (Delgado et al., 1993). One such protein pair is found in the plasmid-borne copper resistance system, encoded by the *pco* operon (Brown et al., 1992). A similar system is encoded by the *Pseudomonas syringae* copper resistance *cop* operon. The two-component regulatory system genes *pcoRS* and *copRS* are homologous, but little is known about their DNA or metal binding properties (Silver and Ji, 1994). Both components of each system are constitutively expressed. It has been suggested, based on the roles of EnvZ and OmpR and sequence homologies with them, that CopS is a transmembrane protein that senses the levels of copper in the periplasm and phosphorylates CopR in response to high copper concentrations (Mills et al., 1993). CopR would thus be activated to induce transcription of the *cop* operon.

A different system of copper-resistance is utilized in *Enterococcus hirae* (Odermatt and Solioz, 1995). The CopY and CopZ proteins regulate a system of copper

uptake and extrusion mediated by the CopA and CopB proteins. While not fully understood, it appears that CopY is the DNA binding repressor protein and CopZ activates transcription of *copAB* when copper levels are high by inducing CopY to dissociate from operator DNA. It has been proposed that CopY must also bind copper when copper levels are low in order to bind DNA, providing a tight range of copper concentrations that would allow repression of the *copAB* genes.

### The MerR family

The MerR family of transcriptional regulators is a group of proteins named after the first member described, the mercuric resistance regulator (Summers and Silver, 1972). These proteins regulate transcription of various operons in response to specific cellular stresses, whether they be toxic levels of heavy metal ions (MerR, ZntR, CueR, etc.), oxygen radicals (SoxR), or a variety of drugs and antibiotics (BmrR, TipA) (reviewed in (Summers, 1992)). Most, if not all, MerR family members bind DNA as homodimers regardless of whether ligand is present, repressing transcription in the absence and activating it in the presence of ligand. These proteins share conserved N-terminal DNA binding domains and have divergent C-terminal effector binding domains. To date, five MerR family members have been characterized that respond to heavy metal ions: MerR [Hg(II)], PbrR [Pb(II)], ZntR [Zn(II), Cd(II), Pb(II)], CoaR [Co(II)], and CueR [Cu(I), Ag(I)] (Stoyanov et al., 2001). *E. coli* SoxR uses two [2Fe-2S] clusters per homodimer to sense changes in oxidation levels in cells, but the FeS cluster is always bound, unlike the heavy metal-responsive MerR family members that bind metals only when they are near toxic levels (Gaudu and Weiss, 1996). TipA binds thiostrepton,



BmrR binds lipophilic cationic compounds, and BltR and Mta ligands have eluded identification but are dissimilar to those that bind BmrR (Ahmed et al., 1995; Baranova et al., 1999). Many other putative MerR family members have been found as a result of bacterial genome sequencing, but assignment of ligand specificity to these hypothetical proteins has been slow.

Because MerR is the most extensively studied member of the family, I shall draw on the wealth of information available for MerR to describe the general mechanism of the family. All family members characterized to date appear to have similar modes of action (for examples, see (Brocklehurst et al., 1999; Stoyanov et al., 2001)). MerR family members have very similar DNA binding domains (Figure 6.8B), and it is likely that they all bind DNA in an analogous manner. The mechanism of repression/activation used by the MerR family is unique.

MerR is bound to its operator DNA at all times. In the absence of its effector, Hg(II), MerR represses transcription from the *mer* operon. MerR also represses transcription of its own divergently transcribed gene, *merR*, in the presence or absence of Hg(II). When Hg(II) levels increase, MerR activates the transcription of the *mer* operon. This operon usually includes genes *merTPABD*. MerT, P, and A transport Hg(II) across the cell membrane to MerB, an organomercurial lyase, and MerA, involved in the reduction of Hg(II) to volatile Hg(0) (Misra, 1992). If present, MerD is a secondary regulatory protein that binds the same DNA site as MerR. The  $K_D$  for MerR binding Hg(II) is about 20 nM. This is a very high affinity for metal and likely reflects the relative toxicity of Hg(II). The residues that bind mercuric ion are Cys79 from one monomer and Cys114 and Cys123 from the other monomer (Helmann et al., 1990). The

Hg is tricoordinated by these cysteines in a trigonal planar configuration (Utschig et al., 1995).

MerR activates transcription in the presence of Hg and represses transcription in its absence by distorting the DNA differentially depending upon the presence of ligand, as depicted in Figure 1.15. MerR, and all MerR proteins that have been thus characterized, binds to a promoter that has suboptimal spacing of 18-20 bp between its -35 and -10 elements. The 17 bp spacer found in most promoters places the -35 and -10 elements in phase on the DNA, thus facilitating recognition by RNAP. A 19 bp spacer between the -35 and -10 elements puts the elements "out of phase" with one another, assuming the DNA is B-form, placing them 70° out of register when looking down the DNA helical axis (Figure 1.7B). The optimal spacing for recognition by *E. coli* RNAP is 17 bp (Jacquet and Reiss, 1990). In fact, if 2 bp are deleted from the spacer region, transcription from the *mer* promoter is constitutive and very strong (Parkhill and Brown, 1990). Adding Hg-MerR to that system decreases the activity of the constitutive promoter, implying that Hg-MerR effectively shortens the DNA.

RNAP is trapped at MerR-DNA complexes even without Hg(II) present. Footprinting studies have shown that RNAP is located just upstream of MerR, contacting the -35 element in the absence of Hg(II) (O'Halloran et al., 1989). Once Hg(II) binds, RNAP extends its contacts to the -10 element and further downstream, locally melting the DNA and thus forming the transcriptionally active open complex. Therefore, MerR actually regulates transcription initiation at the isomerization step, allowing transition from the closed to the open complex in response to Hg(II) binding by altering the conformation of the promoter.

The DNA is distorted by MerR, and this distortion is altered by the addition of Hg(II). Footprinting studies with MerR have shown an enhanced reactivity of the Hg-MerR-DNA complex to intercalating chemical nucleases, such as iron-methidiumpropyl-EDTA (MPE), which suggests that an unwinding locus is at the center of the operator palindrome (Frantz and O'Halloran, 1990). The hypersensitivity to cleavage of DNA at the center, and therefore the distortion of the DNA, correlates with an increase in the steady-state levels of transcription. Other experiments showed that MerR-induced DNA underwinding in the absence of Hg was about 19° (Ansari et al., 1992). When Hg was added the underwinding increased to 52°.

#### The BmrR-drug-DNA structure

The three-dimensional structure of MerR is not known. However, the first structure of a family member, BmrR, bound to DNA and a small molecule coactivator, was recently determined (Heldwein and Brennan, 2001). BmrR is a multidrug binding protein that has the characteristic N-terminal DNA binding domain of other MerR family proteins, but a divergent drug binding C-terminal domain, which is much larger than the C-terminal domains found in the metal binding members of the family (Ahmed et al., 1994). The BmrR-drug-DNA structure showed for the first time the atomic details of the activating DNA distortion mechanism. The 19 bp spacer *is* effectively shortened. The AT bp at the pseudodyad of the operator breaks its Watson-Crick hydrogen bonds, and the adenine and thymidine slide away from each other in the 3' direction. The operator "bunches up" in the middle and is bent by ~50° toward the major groove, away from the protein. The untwisting and base sliding shorten the promoter by ~5 Å, bringing the -35

and -10 elements into phase. A key remaining question is what is the structural mechanism of repression by MerR family members when the "coactivators" are not present. The apo-BmrR-DNA crystal structure would tell us the mechanism of repression in atomic detail. The DNA is distorted somewhat in the repressed form; it is still underwound 19°, and most MerR family proteins do repress transcription levels two- to four-fold below basal levels measured in the absence of protein.

#### The *Proteus mirabilis* transcription regulator, PMTR

DNA binding mechanisms within the MerR family appear to be very similar, based on available biochemical data (for example, (Baranova et al., 1999; Hidalgo et al., 1998; Outten et al., 1999; Outten et al., 2000; Summers, 1992)). DNA distortion analyses by footprinting experiments have all yielded similar results (Baranova et al., 1999; Frantz and O'Halloran, 1990; Outten et al., 1999; Outten et al., 2000). My own results with the *Proteus mirabilis* heavy metal-dependent transcription regulator PMTR are also consistent with those reported previously and are discussed fully in Chapter 6.

PMTR was discovered when the Lutsenko laboratory found that transfection of a genomic fragment from *P. mirabilis* into *E. coli* conferred increased resistance to normally toxic levels of zinc (Noll et al., 1998). Subsequent analysis revealed that PMTR was responsible for the increased amounts and periplasmic accumulation of a 12 kDa proteolytic fragment of an endogenous *E. coli* gene, YJAI. The C-terminal fragment of YJAI bound Zn(II) specifically, but displayed some affinity for Ni(II) or Co(II). The role of YJAI in zinc resistance is not known, but it may modulate the metal specificity of the P-type ATPase ZntA, which transports Pb(II), Cd(II), and Zn(II). While it is not

known which operon PMTR induces in response to increased zinc levels in *E. coli*, I shall present data that indicate PMTR can bind *E. coli zntA* promoter DNA specifically and with high affinity (Chapter 6). This promoter, which has a 20 bp spacer, is normally bound by another MerR family member, ZntR, which was discovered shortly after PMTR (Brocklehurst et al., 1999). ZntR and PMTR have high sequence identity (34%), thus supporting their possible overlapping regulatory roles.

To begin to understand the role of PMTR in *P. mirabilis*, we undertook studies to determine its cognate DNA binding site. Electrophoretic mobility shift assays (EMSA) and footprinting experiments revealed that PMTR bound a specific DNA site located immediately upstream of a putative Cu-ATPase on a genomic fragment from *P. mirabilis*. Subtle changes in the footprint were observed upon addition of zinc. Subsequent *in vitro* equilibrium binding experiments showed that PMTR bound this operator with high affinity ( $K_d \approx 20$  nM). No information is available about the function of the putative Cu-ATPase in *P. mirabilis* cells. Furthermore, we do not yet know if other promoters are regulated by PMTR *in vivo*.

PMTR shows high homology with *E. coli* MerR family members, particularly those that bind heavy metals such as ZntR, CueR, and MerR (Figure 6.8B, Table 6.2). The highest homology is with the recently discovered CueR (37% identity), a Cu(I) and Ag(I) binding regulator of *copA*, a CPx-type copper-transporting ATPase (Outten et al., 2000). Two other putative copper binding MerR family members have been reported from *Sinorhizobium meliloti* and *Rhizobium leguminosarum*. Only sequence information is available for these proteins at this time (W. Reeve and M. Dilworth, personal communication). Remarkably, PMTR has 57% and 46% sequence identity with these

heavy metal dependent transcription regulators, or HMDTRs. The residues involved in tricoordination of Hg(II) in MerR are known and all three are cysteines. These cysteines are conserved in *E. coli* ZntR, which likely has at least one more protein ligand contributing specificity for the zinc ion, as zinc prefers to be at least tetracoordinated in proteins (Alberts et al., 1998; Harding, 1999). One of these conserved cysteines (Cys79 in MerR) is a serine in PMTR, CueR, and the two soil bacteria HMDTR proteins. In fact, a short conserved sequence is found across the four proteins instead of the mercury-coordinating Cys79: R-(T/A/H)-S-(A/S)-D-V-(K/R), where the S is in the position of Cys79 that is found in MerR, ZntR, and PbrR. This sequence is not found in any of the other MerR family members. These four proteins also share a shorter 7-residue spacing between the two other cysteines corresponding to the two others that bind Hg in MerR. This spacer is 8 residues in MerR and 9 in ZntR. This difference should alter the relative location of these cysteines, which might facilitate the binding of a different metal.

In conjunction with the DNA binding studies, Dr. Svetlana Lutsenko has determined the *in vitro* metal specificity of PMTR. She found that purified PMTR bound zinc with higher affinity than cadmium or copper, and no binding of cobalt was detected. These studies support the *in vivo* results that led to the discovery of PMTR, as well as the footprinting data that show subtle differences in the presence and absence of zinc. Interestingly, they show that PMTR has different metal specificity from either CueR or ZntR.

Footprinting experiments with CueR have revealed an operator that shares 75% identity with the operator bound by PMTR (Figure 6.8A, Table 6.1). The -35 element and entire left dyad are conserved, with the PMTR operator diverging from a perfect

palindrome in the right dyad and one bp difference between the two -10 elements. This may not seem extraordinary, since all family members bind these suboptimal promoters, but the *E. coli zntA* promoter that ZntR binds has only 15% identity in the same range, most of it in the -35 and -10 elements, which are relatively conserved in all promoters.

It is clear that further experiments are required to determine the function of PMTR in *P. mirabilis*, but we have laid the groundwork for such studies by identifying one promoter/operator that is likely used *in vivo* by PMTR, and experiments to affirm this are ongoing. Future experiments might reveal the role of the *P. mirabilis* putative copper ATPase.

In conclusion, PMTR is a recently identified protein that was thought to be a member of the MerR family of transcription regulators. PMTR was known to modulate resistance to normally toxic levels of zinc. However, nothing was known about the DNA binding properties, such as specificity or affinity, of PMTR. To determine the cognate DNA site for PMTR, we initiated EMSA and footprinting experiments. We found one DNA site that was specifically bound by PMTR in those experiments, which showed subtle differences in footprint upon addition of zinc. Once the specific DNA binding site was determined, the affinity of PMTR for the site was measured. The operator bound by PMTR aligns well with others bound by MerR family members. We have also determined the *in vitro* metal specificity of PMTR, the first such characterization of any MerR family member.

Table 1.1. A taxonomy of prokaryotic transcription regulators for which DNA binding domain structures are available. Proteins are grouped by DNA binding motif. The organisms in which they are found, presence (+) or absence (-) of DNA in structures, and earliest structure citations are given.

Protein	Organism	+/- DNA	Reference
<b>Helix-turn-helix</b>			
cI	Phage $\lambda$	-	(Pabo and Lewis, 1982)
cI	Phage $\lambda$	+	(Jordan and Pabo, 1988)
Cro	Phage $\lambda$	-	(Anderson et al., 1981)
Cro	Phage $\lambda$	+	(Brennan et al., 1990)
Repressor	Phage 434	-	(Mondragon et al., 1989a)
Repressor	Phage 434	+	(Aggarwal et al., 1988)
Cro	Phage 434	-	(Mondragon et al., 1989b)
Cro	Phage 434	+	(Wolberger et al., 1988)
Trp repressor	<i>E. coli</i>	-	(Lawson et al., 1988)
Trp repressor	<i>E. coli</i>	+	(Otwinowski et al., 1988)
Purine repressor	<i>E. coli</i>	-	(Nagadoi et al., 1995)
Purine repressor	<i>E. coli</i>	+	(Schumacher et al., 1994a)
Lac repressor	<i>E. coli</i>	-	(Kaptein et al., 1985)
Lac repressor	<i>E. coli</i>	+	(Lewis et al., 1996)
Fructose repressor	<i>E. coli</i>	-	(Penin et al., 1997)
Fis	<i>E. coli</i>	-	(Kostrewa et al., 1991)
DtxR	<i>C. diphtheriae</i>	-	(Schiering et al., 1995)
DtxR	<i>C. diphtheriae</i>	+	(White et al., 1998)
IdeR	<i>M. tuberculosis</i>	-	(Pohl et al., 1999)
SmtB	<i>Synechococcus</i>	-	(Cook et al., 1998)
MarA	<i>E. coli</i>	+	(Rhee et al., 1998)
Rob	<i>E. coli</i>	+	(Kwon et al., 2000)
NtrC	<i>S. typhimurium</i>	-	(Pelton et al., 1999)
Tet repressor	<i>E. coli</i>	-	(Hinrichs et al., 1994)
Tet repressor	<i>E. coli</i>	+	(Orth et al., 2000)
c2	Phage P22	-	(Sevilla-Sierra et al., 1994)
SinR	<i>B. subtilis</i>	-	(Lewis et al., 1998)
FlhD	<i>E. coli</i>	-	(Campos et al., 2001)
TyrR	<i>H. influenzae</i>	-	(Wang et al., 2001)
<b>Winged helix</b>			
CRP	<i>E. coli</i>	-	(McKay et al., 1982)
CRP	<i>E. coli</i>	+	(Schultz et al., 1991)
BirA	<i>E. coli</i>	-	(Wilson et al., 1992)
OmpR	<i>E. coli</i>	-	(Martinez-Hackert and Stock, 1997)
PhoB	<i>E. coli</i>	-	(Okamura et al., 2000)



NarL	<i>E. coli</i>	-	(Baikalov et al., 1996)
MuA	Phage Mu	-	(Clubb et al., 1994)
MuR	Phage Mu	-	(Ilangovan et al., 1999)
MuR	Phage Mu	+	(Wojciak et al., 2001)
BmrR	<i>B. subtilis</i>	+	(Heldwein and Brennan, 2001)
ModE	<i>E. coli</i>	-	(Hall et al., 1999)
CooA	<i>R. rubrum</i>	-	(Lanzilotta et al., 2000)
LexA	<i>E. coli</i>	-	(Fogh et al., 1994)
Arginine repressor	<i>E. coli</i>	-	(Sunnerhagen et al., 1997)
<b>β-ribbon</b>			
MetJ	<i>E. coli</i>	-	(Rafferty et al., 1989)
MetJ	<i>E. coli</i>	+	(Somers and Phillips, 1992)
Mnt	Phage P22	-	(Burgering et al., 1994)
Arc repressor	Phage P22	-	(Breg et al., 1990)
Arc repressor	Phage P22	+	(Schildbach et al., 1999)
CopG	pMV158	+/-	(Gomis-Rüth et al., 1998)
IHF	<i>E. coli</i>	+	(Rice et al., 1996)
HU	<i>B. stearothermophilus</i>	-	(Tanaka et al., 1984)
TFI	Phage SPO1	-	(Jia et al., 1996)

Table 1.2. Alignment of  $O_R$  and  $O_L$  sites. Below the consensus, the number of occurrences and position numbers are given. Adapted from (Albright and Matthews, 1998b).

Operator	Sequence	
	Consensus half	Nonconsensus half
$O_R1$	5' T A T C A C C G C C A G A G G T A 3' A T A G T G G C G G T C T C C A T	
$O_R2$	5' T A A C A C C G T G C G T G T T G 3' A T T G T G G C A C G C A C A A C	
$O_R3$	5' T A T C A C C G C A A G G G A T A 3' A T A G T G G C G T T C C C T A T	
$O_L1$	5' T A T C A C C G C C A G T G G T A 3' A T A G T G G C G G T C A C C A T	
$O_L2$	5' C A A C A C C G C C A G A G A T A 3' G T T G T G G C G G T C T C T A T	
$O_L3$	5' T A T C A C C G C A G A T G G T T 3' A T A G T G G C G T C T A C C A A	
Consensus	5' T A T C A C C G C C A G T G G T A 3' A T A G T G G C G G T C A C C A T	
# of occurrences	5 6 4 6 6 6 6 6 5 3 4 5 3 6 3 6 4	
Position	1 2 3 4 5 6 7 8 9 8'7'6'5'4'3'2'1'	

Figure 1.1A. The transcription initiation complex. RNA polymerase holoenzyme is shown bound to a promoter containing the UP, -35, and -10 elements. B. The four steps of transcription initiation by *E. coli* RNA polymerase. Adapted from (Herschbach and Johnson, 1993).

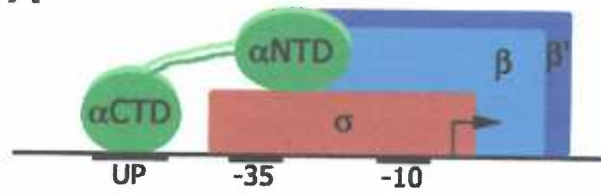
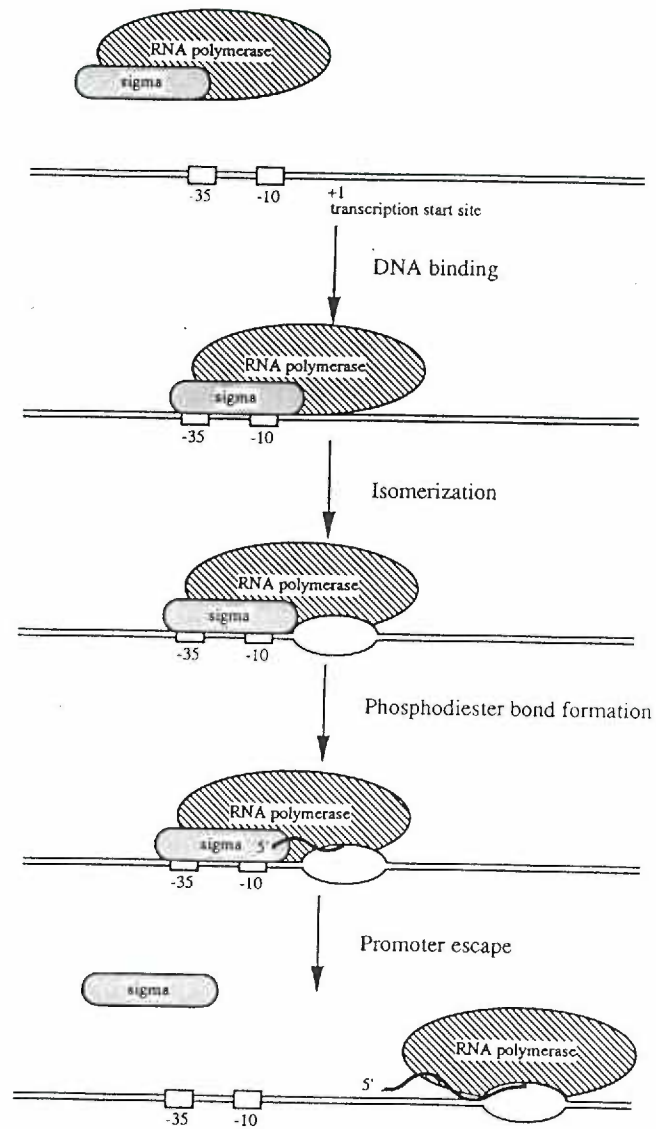
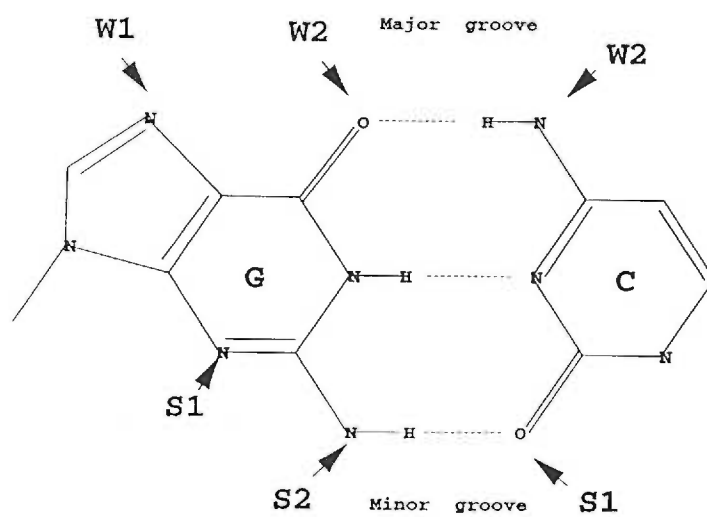
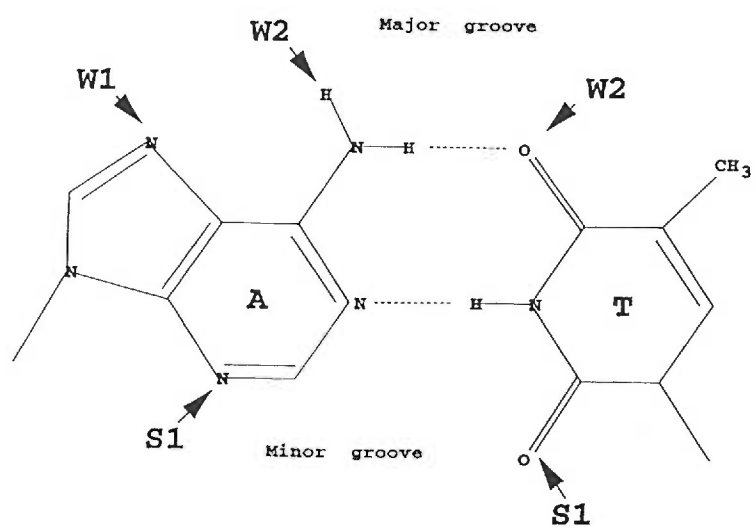
**A****B**

Figure 1.2A. Watson-Crick base pairs as found in B-DNA. Adapted from (Mandel-Gutfreund et al., 1995). W1, W2, S1, and S2 mark positions available for hydrogen bonding as determined by (Seeman et al., 1976). B. Local DNA helix parameters for pairs of adjacent base pairs. From (Dickerson, 1998).

**A**



# B

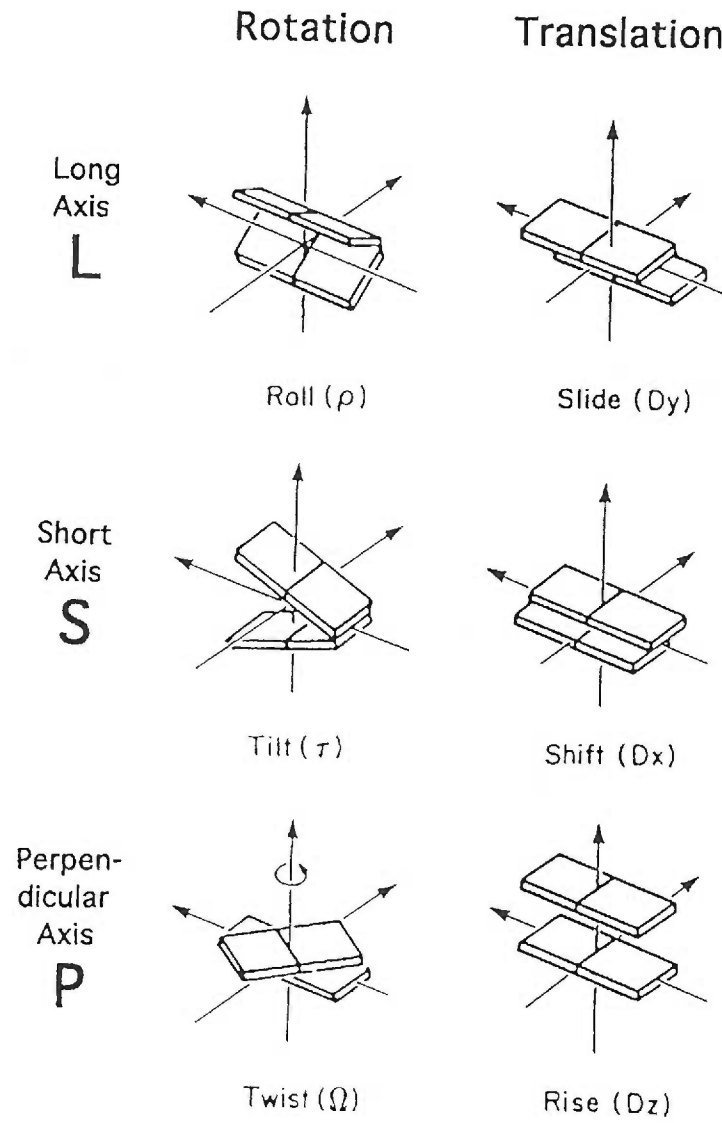
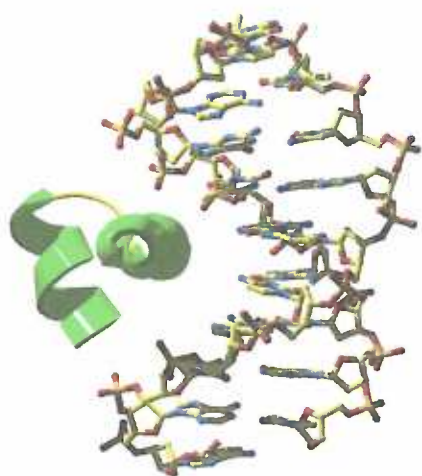


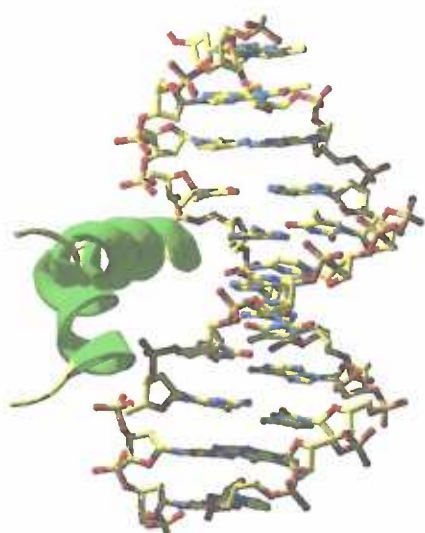
Figure 1.3. Helix-turn-helix motifs from three proteins bound to DNA. A.  $\lambda$  cI (PDB code 1lmb) B. Trp repressor (PDB code 1trr). C.  $\lambda$  Cro (PDB code 6cro). Recognition helices are near perpendicular to the page. Helices are colored green. Turns or loops are colored yellow. The DNA atoms and bonds are represented as balls and sticks. Carbons are colored yellow, oxygens are red, nitrogens are blue, and phosphates are orange. This color scheme will be used throughout this work.



**A**



**B**



**C**

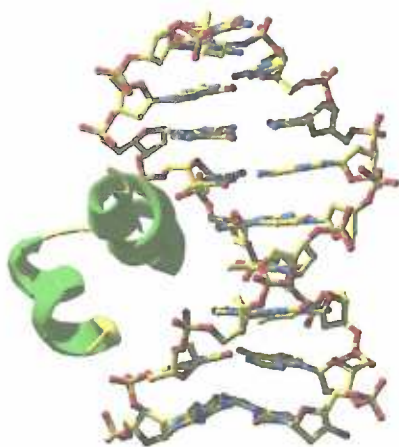


Figure 1.4A.  $\lambda$  cI bound to O<sub>L</sub>1 DNA (PDB code 1lmb). Helices are colored green,  $\beta$ -strands are blue, and loops are yellow. This scheme will be used for all illustrations of proteins in this thesis, unless otherwise noted. B.  $\lambda$  Cro bound to pseudo-consensus DNA (PDB code 6cro). Unless otherwise noted, all illustrations of macromolecules were made using the molecular graphics program SwissPDB Viewer.

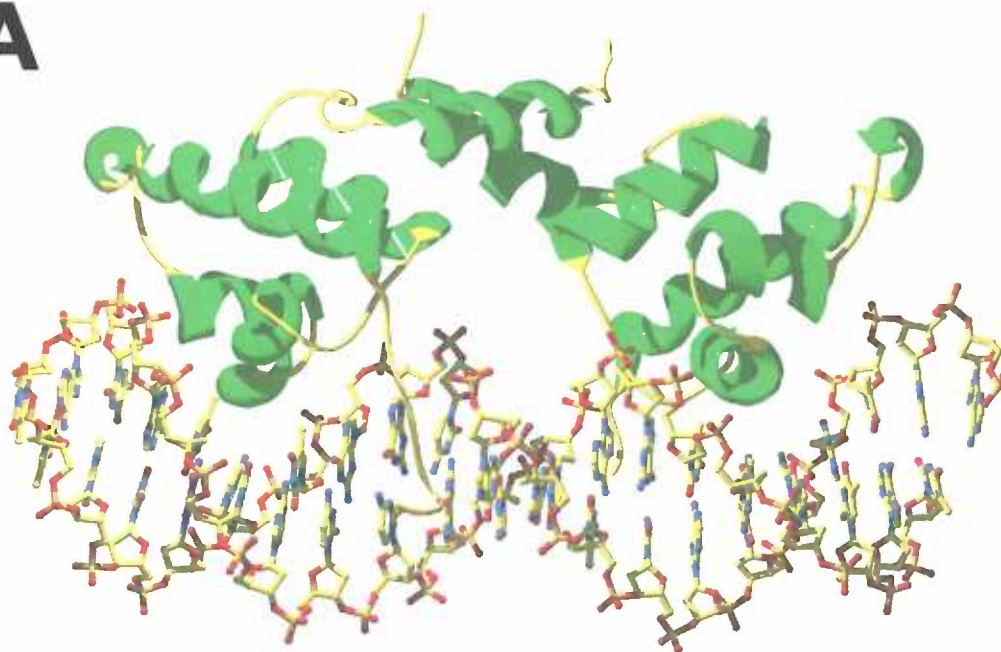
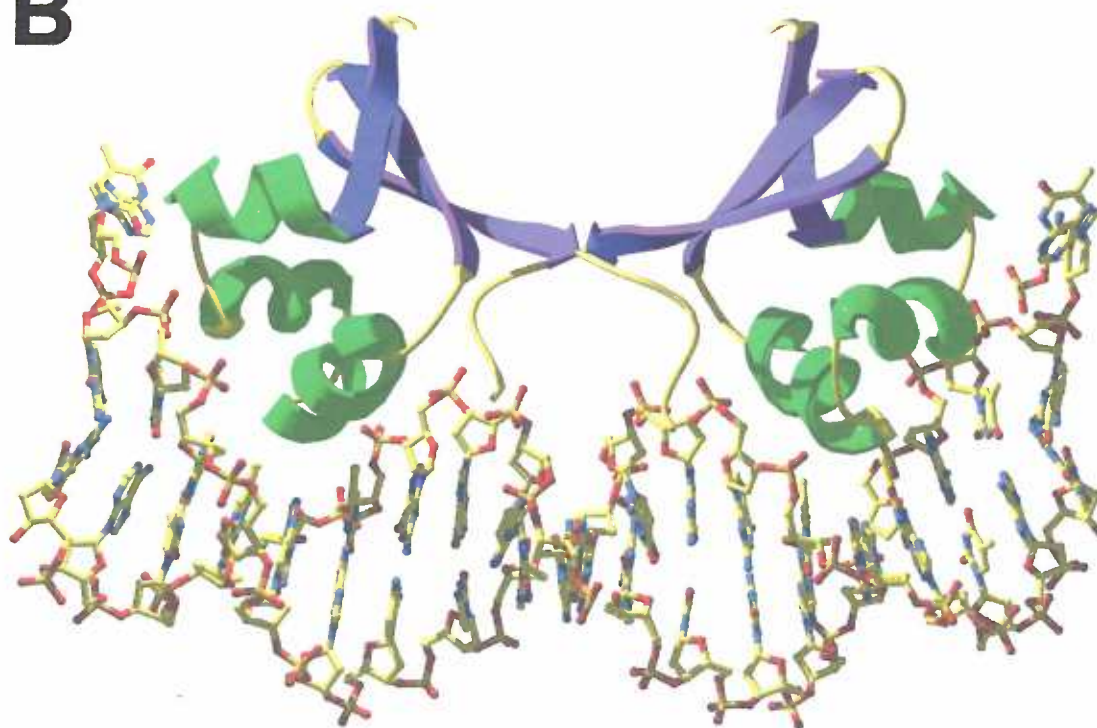
**A****B**

Figure 1.5. Close-up view of CRP bound to DNA, showing only the DNA binding domain of CRP (PDB code 1ber). Three residues, Arg180, Glu181, and Arg185, make extensive contacts with DNA bases.

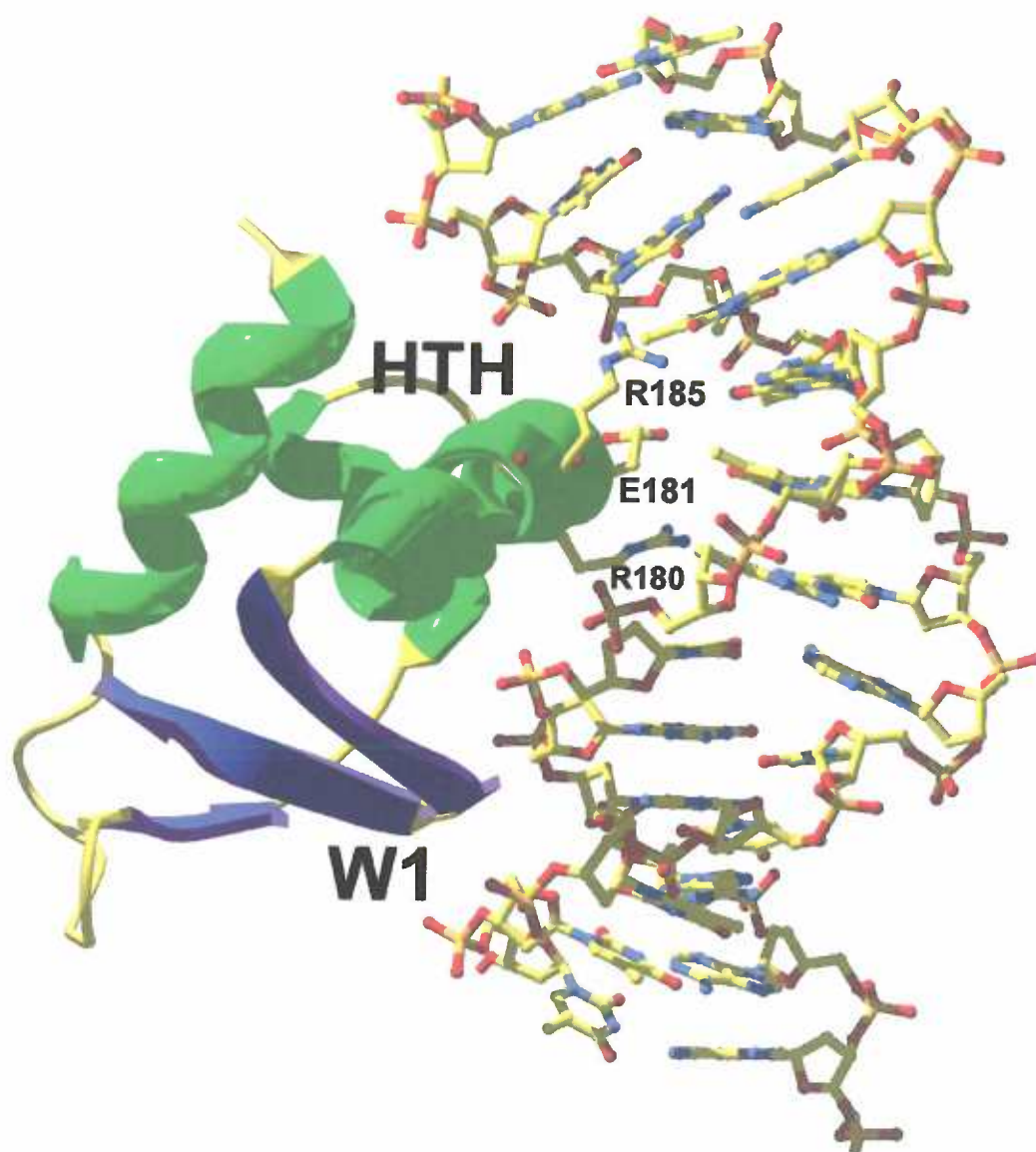
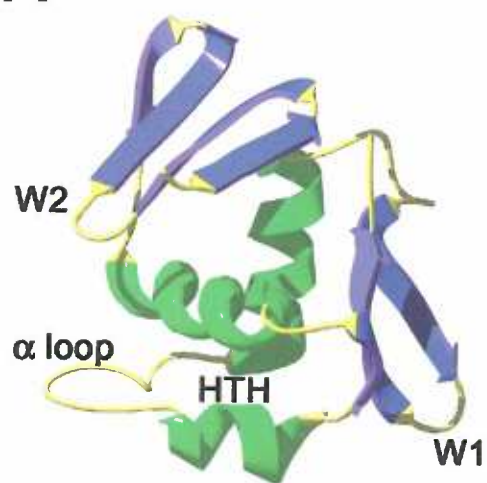
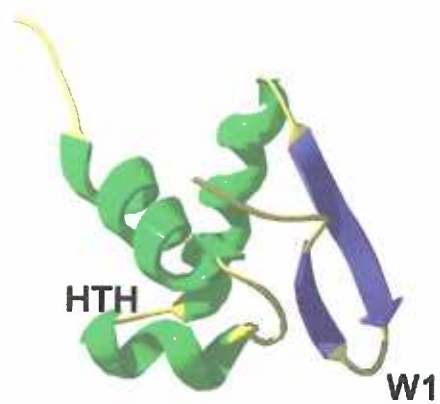


Figure 1.6. Winged helix domains from four proteins: A. OmpR (PDB code 1opc). B. BirA (PDB code 1bia). C. BmrR (PDB code 1exi). D. MuR (PDB code 1g4d). The structures of OmpR and BirA were solved in the absence of DNA. BmrR and MuR structures were determined in the presence of DNA.

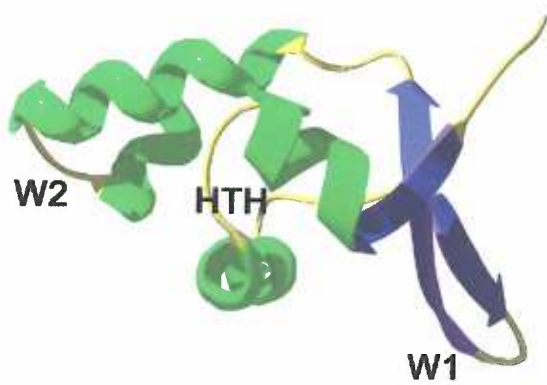
**A**



**B**



**C**



**D**

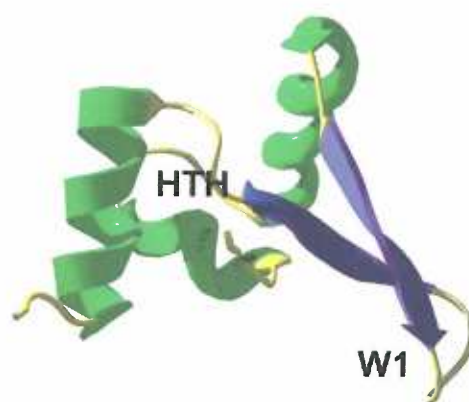


Figure 1.7A. The BmrR-DNA-drug complex (PDB code 1exi). In one BmrR monomer, the DNA binding domain is colored yellow, the linker region is red, and the drug binding domain is green. The other BmrR monomer is colored blue. The DNA and drug molecules are represented as balls and sticks. B. *Top*: B-form DNA in a promoter with a 19 bp spacer. The -35 element is colored pink and the -10 element is colored green. *Middle*: the shape of the DNA in the BmrR-DNA-drug complex, as envisioned in a longer piece of DNA. *Bottom*: B-form DNA in a promoter with a 17 bp spacer. These figures were taken from (Heldwein and Brennan, 2001).



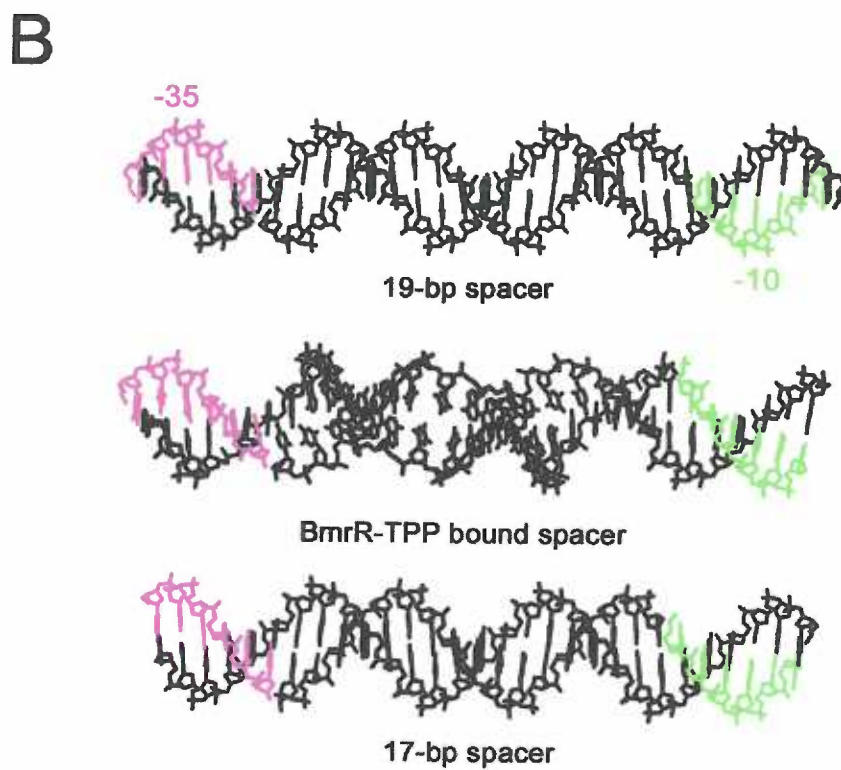
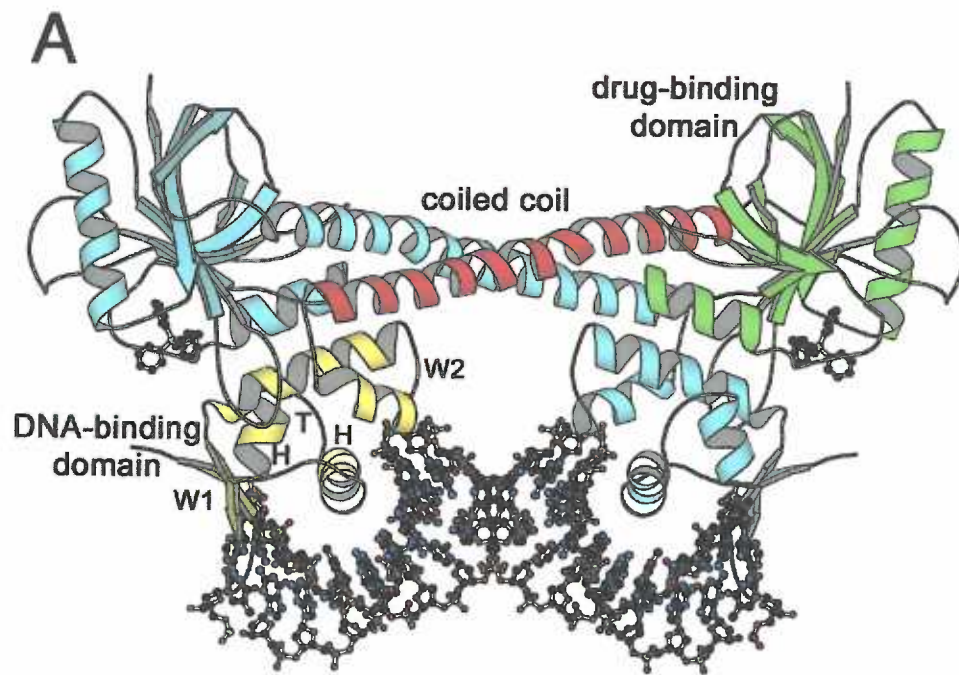


Figure 1.8A. MetJ bound to S-adenosyl methionine, in pink, and DNA (PDB code 1cma). The illustration on the left is a "side-on" view and can be rotated 90° toward the reader to generate the "top-down" view shown on the right. B. The Arc repressor bound to DNA in a "dimer of dimers" configuration (PDB code 1bdt).

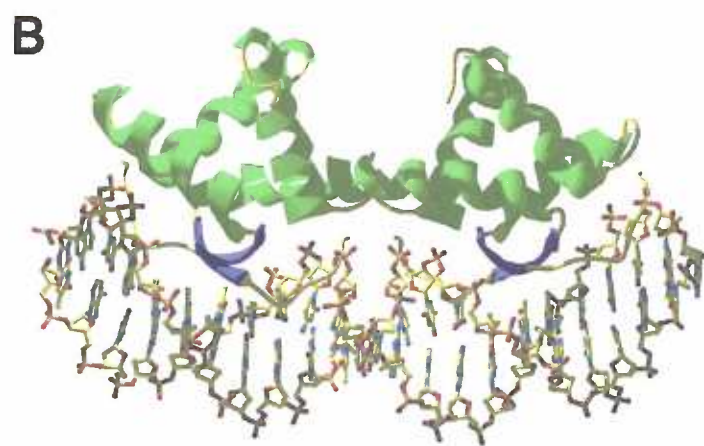
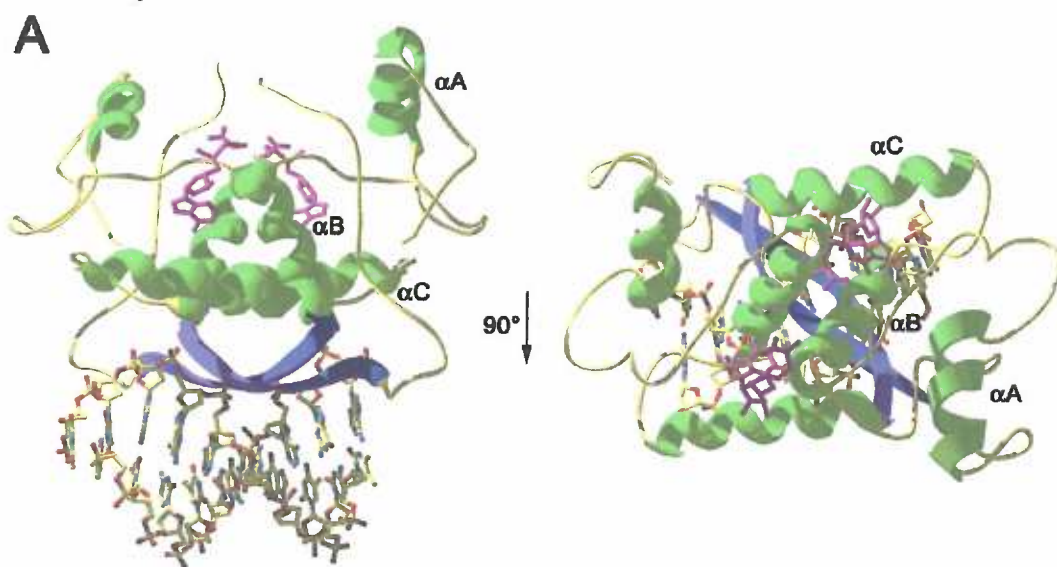


Figure 1.9. Integration host factor bound to 35 bp of DNA (PDB code 1ihf).

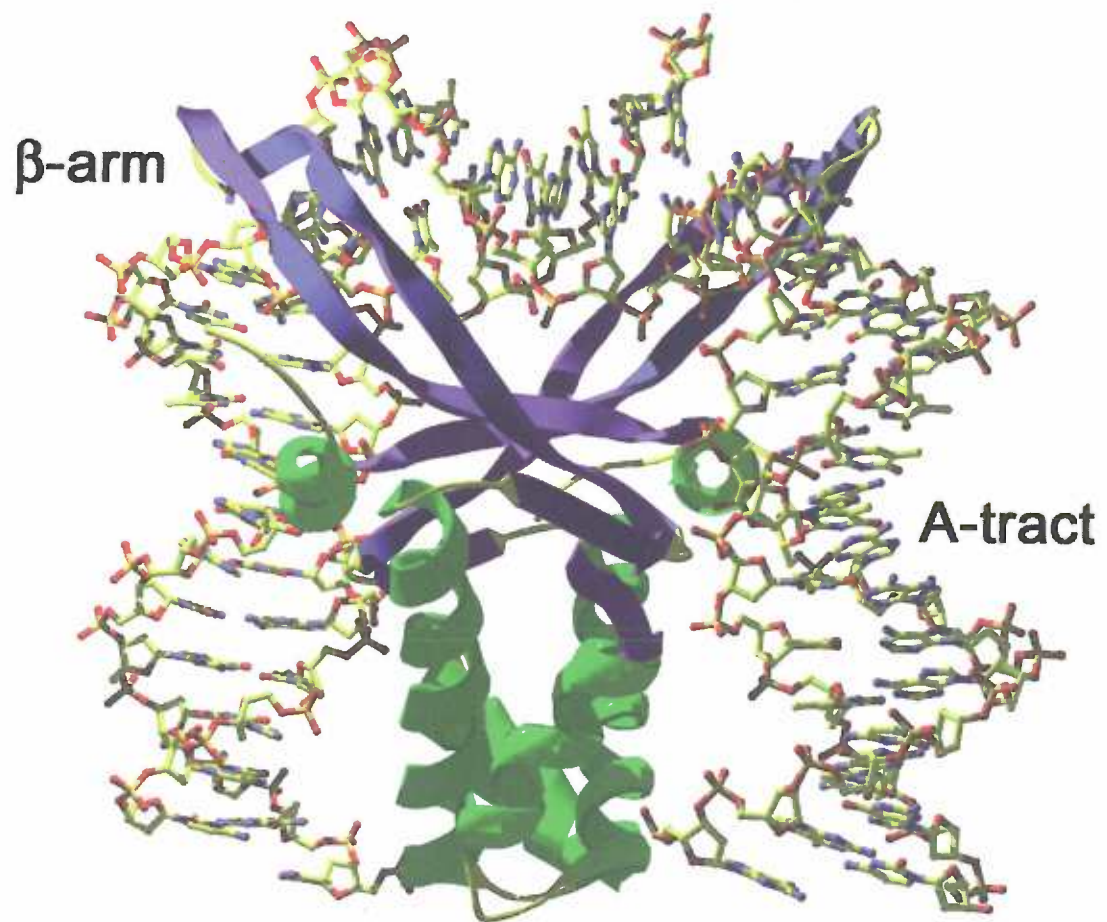


Figure 1.10A. Genes regulated by PurR. B. sequence of *pur* regulon operators recognized by PurR. C. The loci of each PurR operator site depicted as thick black lines. The transcription start sites are marked +1, and translation start sites are shown as thick white boxes. Fold regulation by PurR at each promoter is reported in the right column. Adapted from (Zalkin and Nygaard, 1996).

A

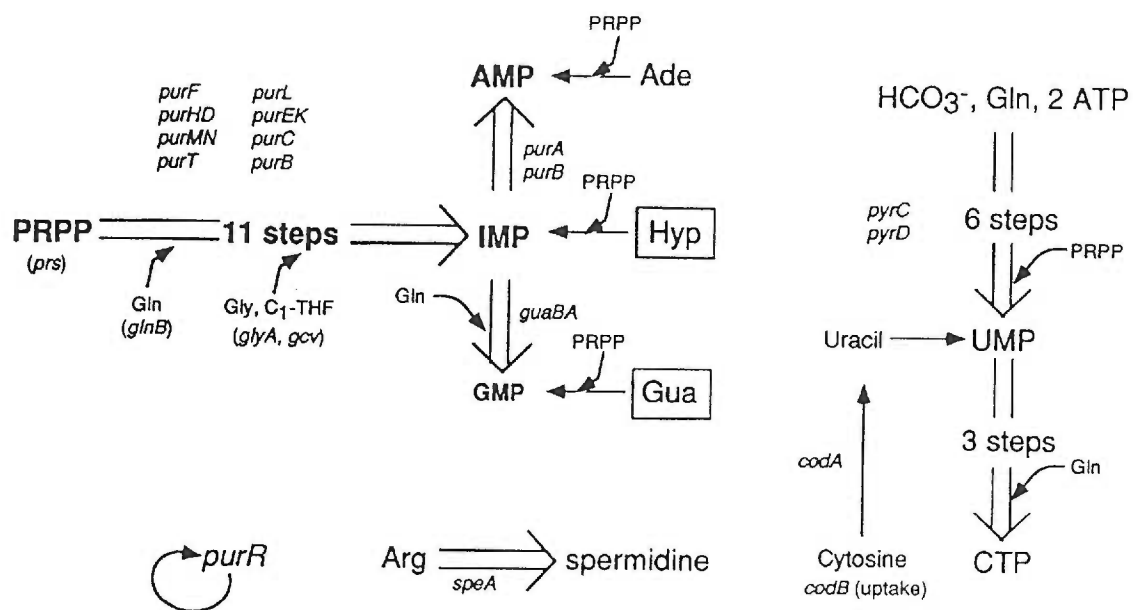






Figure 1.11A. The PurR-hypoxanthine-*purF* operator complex (PDB code 1pnr). The PurR homodimer is shown in green. Hypoxanthine (blue) and DNA (red) are shown as space-filling atomic models. B. The PurR corepressor binding domain, in green (PDB code 1dbq).

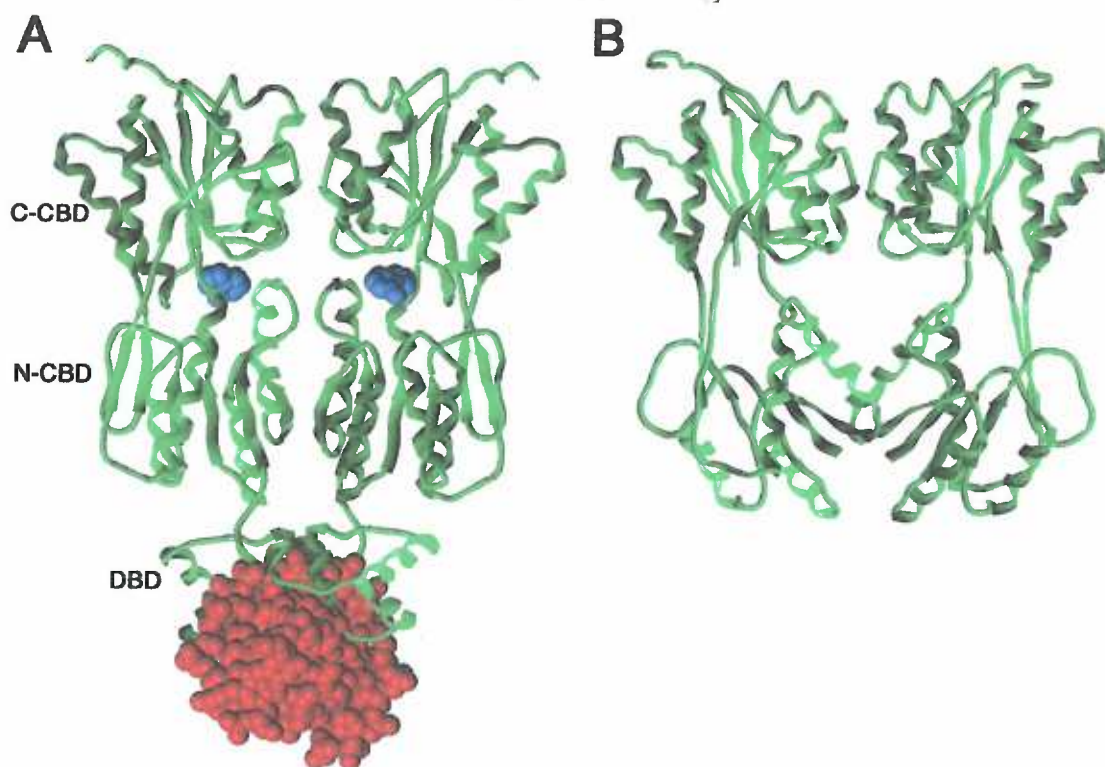


Figure 1.12. Illustrations of the effector binding domains of monomers of LacI/GalR family members and a periplasmic binding protein, highlighting the structural homologies. A. PurR (PDB code 1qpz). B. Ribose binding protein (PDB code 2dri). C. LacI (PDB code 1efa). D. Trehalose repressor (PDB code 1byk). In each model, the N-subdomains are at the bottom and the C-subdomains are at the top.

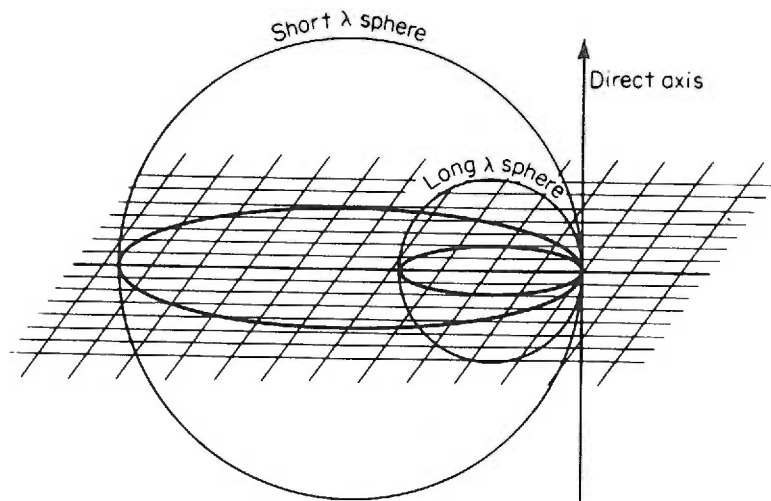
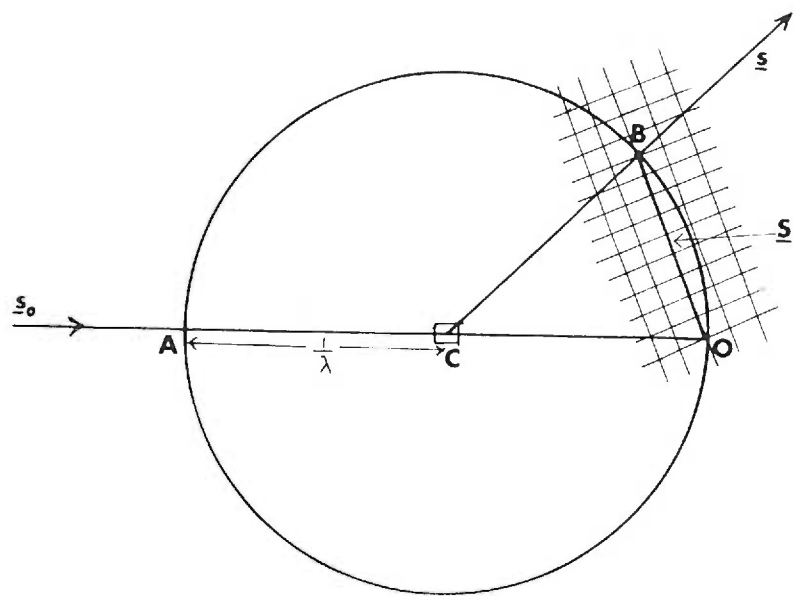


Figure 2.10. Top: The Ewald construction. Modelled after (Blundell and Johnson, 1976). Bottom: The spheres of reflection for two wavelengths. From (Stout and Jensen, 1989).

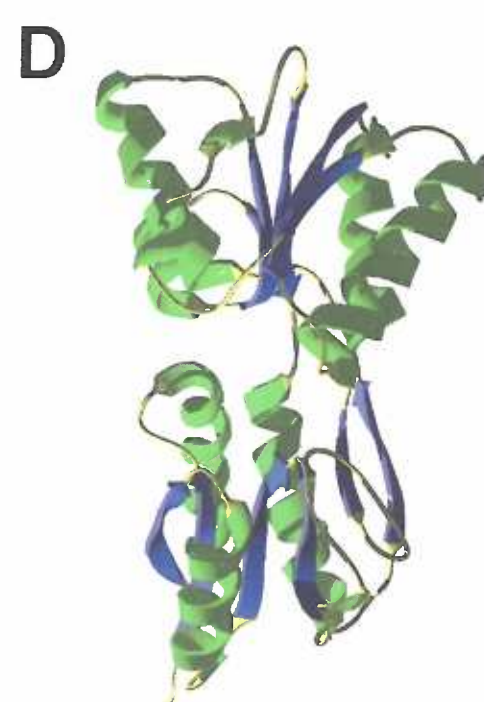
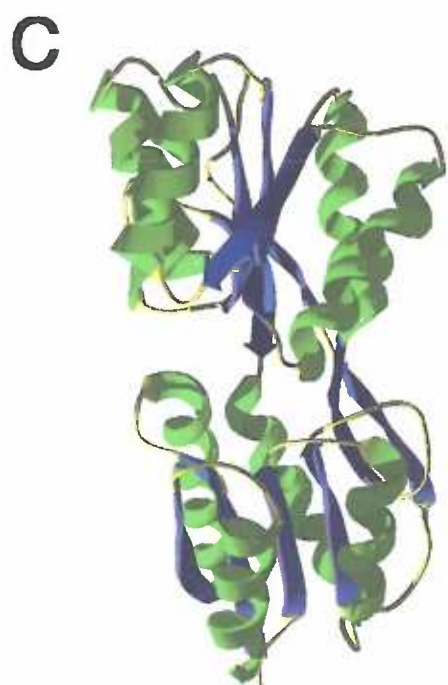
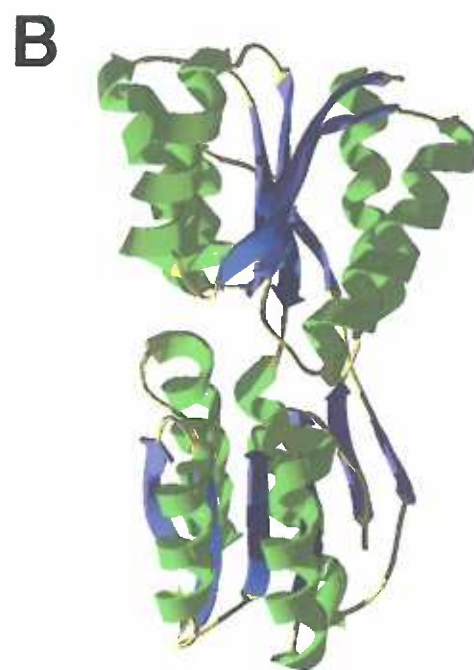
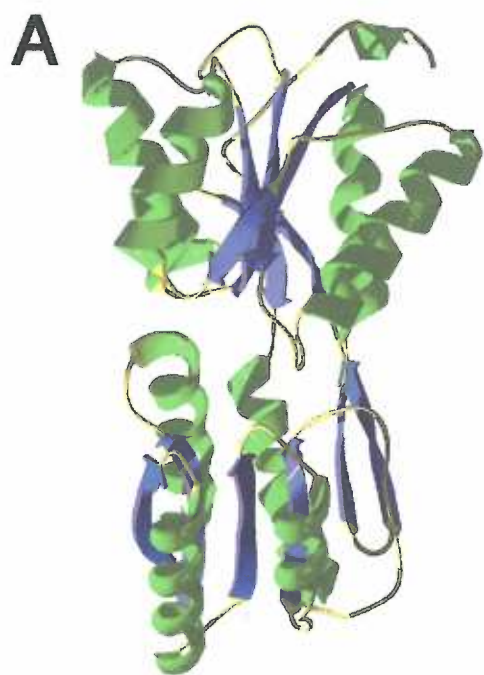


Figure 1.13. Close-up views of the PurR corepressor binding pocket in the apo and holo structures. The protein backbone is depicted as a green worm. A. The hypoxanthine-bound structure. Residues that contact corepressor directly are shown and labelled, as are Trp147 and Tyr126. B. The apo corepressor binding pocket. The same residues are shown as in A, except Arg190 and Thr192, which are disordered in the CBD structure.

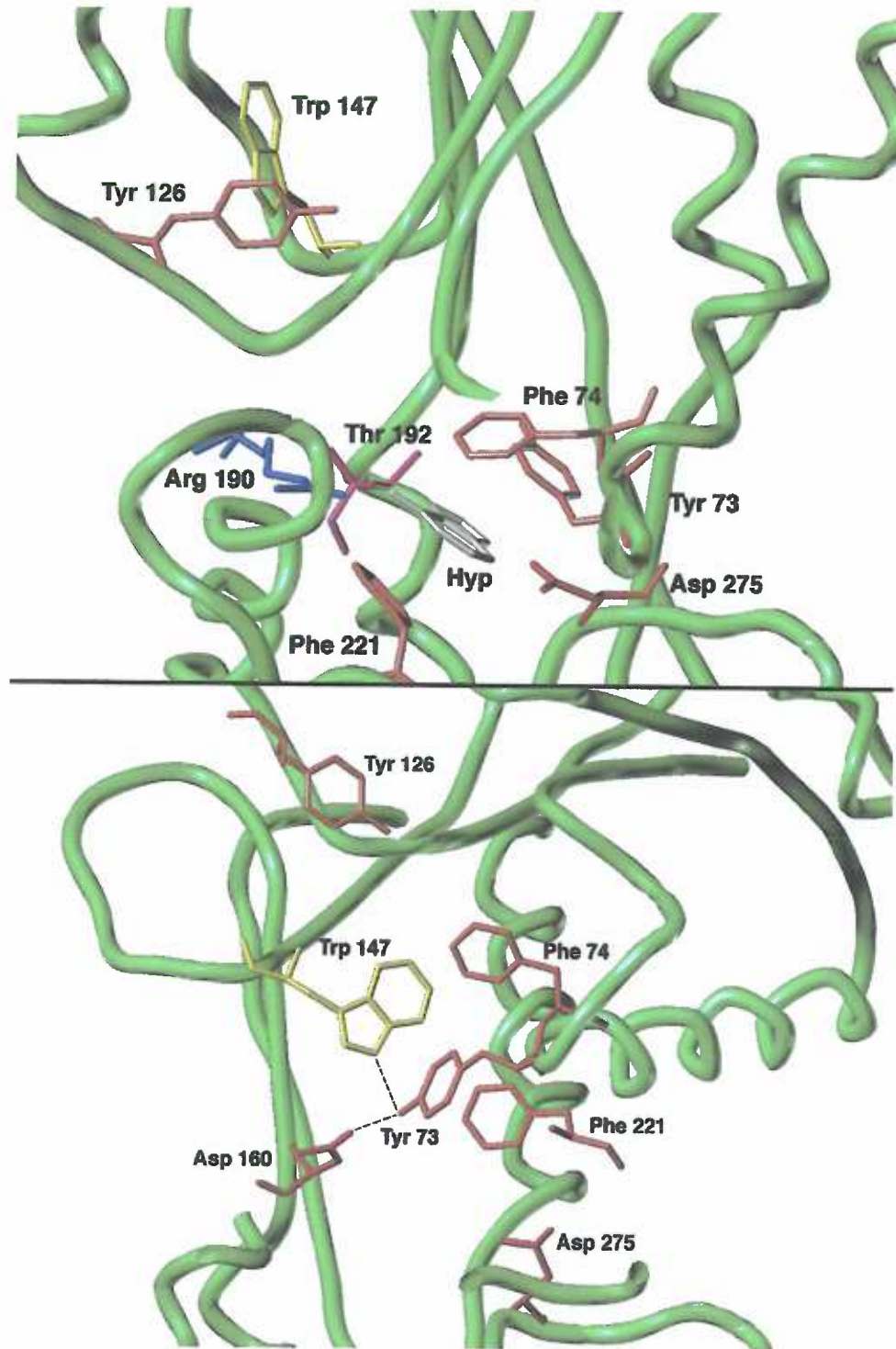


Figure 1.14A. The structure of SmtB (PDB code 1smt). B. Two dimers of DtxR bound to DNA (PDB code 1ddn).



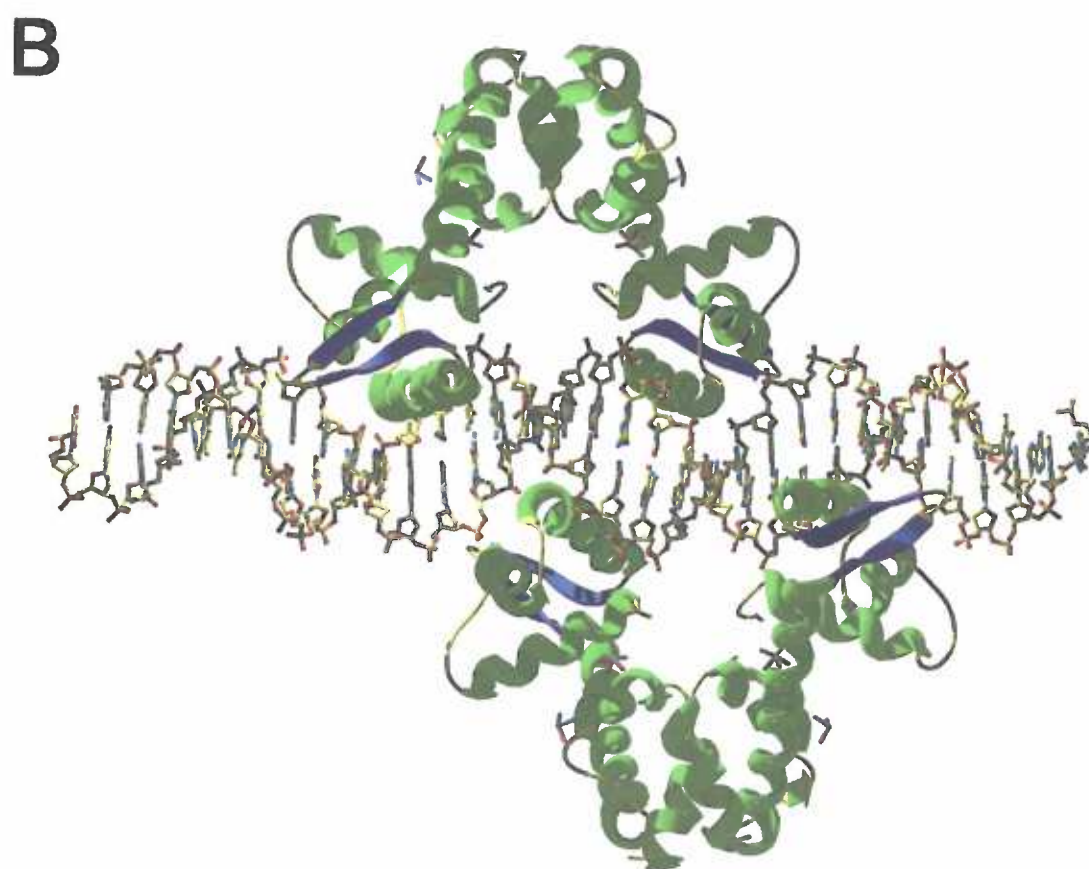
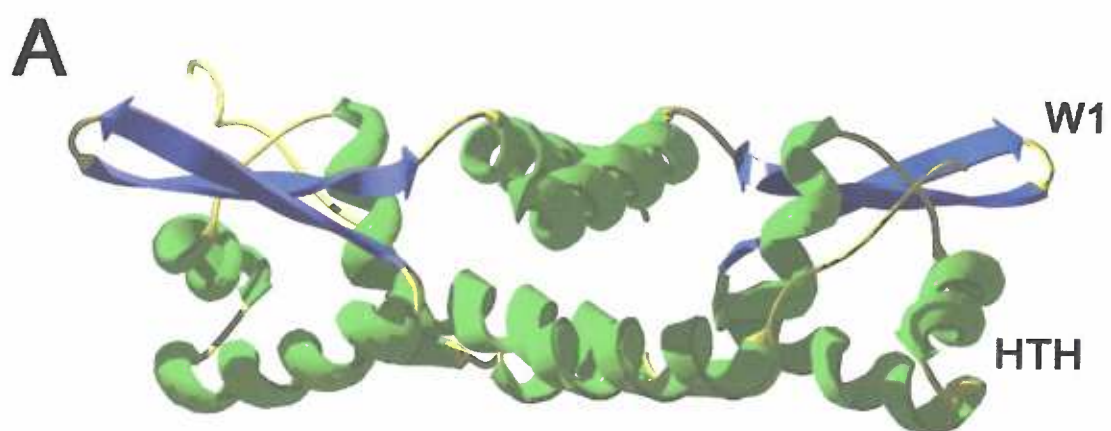
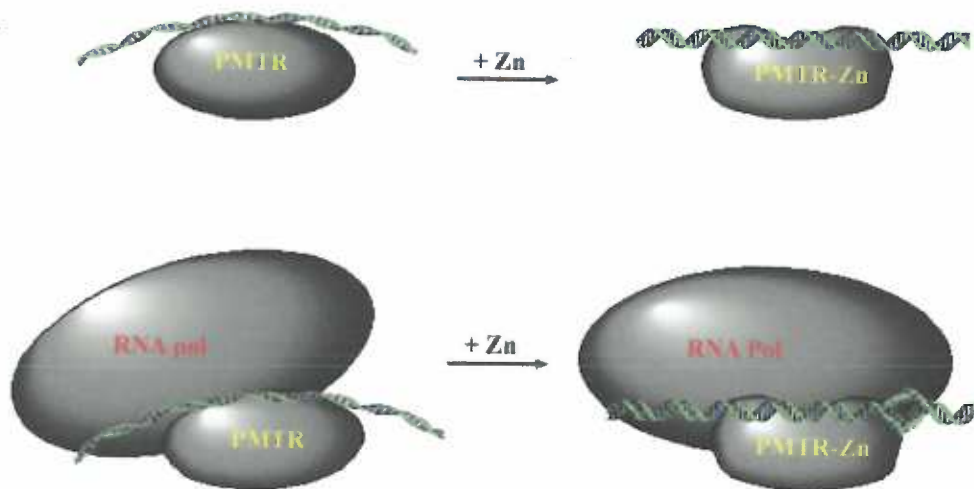


Figure 1.15. The interactions of PMTR with DNA and RNA polymerase. After a model proposed for MerR (Ansari et al., 1995).

## Model of PMTR-DNA interactions



## **Chapter 2. Determination of macromolecular structures using X-ray crystallography**

X-ray crystallography is a powerful tool that is, in reality, a science unto itself. This method allows one to visualize the precise three-dimensional positions of most or all of the atoms in a given molecule. It is inherently multidisciplinary, drawing upon knowledge of chemistry, physics, mathematics, and in its application biology or geology. X-rays were discovered in 1895 by Röntgen, but it was not until 1912 that Max von Laue discovered that X-rays are diffracted by crystals (Friedrich et al., 1912). Around the same time, W.L. Bragg noticed that this diffraction could be treated as "reflection" from planes in the crystal lattice (Bragg, 1913). From these observations, the field of X-ray crystallography was born.

In the early stages, mineral crystals were the most widely studied, partly due to greater availability. However, proteins were already known to crystallize. Protein crystals were first described by Hünefeld in 1840 for hemoglobin, and crystals of several globulins and albumins were obtained in the 19th century. However, these crystals were used for protein purification, not structure determination (described in (McPherson, 1999)). It was not until the 1930s that X-ray diffractionists first experimented with macromolecular crystals, those of proteins and viral particles. The first high resolution structure was not solved until many years later, as new methods needed to be developed for the determination of the thousands of atomic positions in one "small" protein, such as myoglobin (Kendrew et al., 1960). Computing power has improved such that currently, crystal structures can sometimes be solved in a matter of days. However, not all

structures can be determined that quickly, as numerous problems can be encountered along the way.

There are five major steps in determining a structure by X-ray crystallography:

1. Crystallization of the molecule(s) of interest
2. X-ray intensity data collection
3. X-ray intensity data processing
4. Phase determination
5. Interpretation of electron density maps and model refinement

I shall not attempt to cover any of these topics in full. Many books have been written on both crystallization and crystallography. At least five journals specific to X-ray crystallography publish new methods, techniques, theories, and results monthly. Instead, I shall describe the topics that have related most closely to my own research. I will thus try to combine theoretical aspects in a practical way, moving chronologically as one would solve a structure by molecular modification techniques. Within Chapters 3-5 of this dissertation, I report eight structures that I solved by this method. I also report a total of six structures solved by Drs. Dennis Arvidson and Maria Schumacher in Chapter 4.

### Crystallization

Obtaining data quality crystals is clearly essential to the method of X-ray crystallography. This is often the rate-limiting step in a given project. Obtaining data-quality crystals is also one of the most exciting parts of the process. When a young graduate student sees her first crystals under the microscope, she might naively believe that the rest of her graduate career will simply fall into place after this landmark

discovery. However, she soon learns the trials of crystallization: salt crystals are much easier to grow than macromolecular ones; small, overnucleated drops are easier to generate than those with large, single crystals; and physically beautiful, large crystals do not necessarily diffract X-rays well.

Many factors play a role in the probability that a given macromolecule, or macromolecular complex, will crystallize. The more homogeneous and monodisperse the starting material, the more likely it is that it will produce crystals (McPherson, 1999). For purposes of this discussion, we will assume that very high quality starting material is in hand.

Crystallization of any molecule or compound is dependent upon reaching supersaturation in solution. The solubility of most molecules depends on parameters such as ionic strength, pH, and temperature of a solution. There is always a limit, the saturation point, for any specific condition. At the saturation point, there is no net energetic gain from moving to either the solid or solution state. Here, crystallization is not favored (Figure 2.1). When the concentration of solute increases above saturation, that solution is supersaturated. A supersaturated solution is not at equilibrium, but the laws of thermodynamics insist that it move toward equilibrium by excluding some molecules from solution into the solid state. The activation energy barrier to initiation of the solid state is overcome by statistical fluctuations that result in the formation of critical nuclei. The higher the supersaturation, the greater the tendency for molecules to enter an aggregated, solid state. Of course, not only will the formation of crystal nuclei be favored, so will the formation of amorphous precipitate. At very high supersaturation

levels, a shower of microcrystals may result. This phenomenon is termed "overnucleation".

In the metastable state, a few nuclei may form, and the tendency will be for these crystals to grow, effectively lowering the level of supersaturation by removal of molecules into the solid phase (Figure 2.1). This state favors slow growth of the few crystals in the solution. Therefore, the goal of the macromolecular crystallographer is to nucleate just a few crystals, preferably within the labile region, and slowly enter the metastable region. This way the crystals will continue to grow slowly and with fewer defects, such as microheterogeneity or incorporation of impurities.

The preferred method of crystallization in the Brennan laboratory is that of hanging-drop vapor diffusion, depicted in Figure 2.2. Other methods are available, but most rely on the same principles. In this method, a few microliters of macromolecule solution at high concentration (10-20 mg/ml) are mixed with an equal amount of mother liquor—the precipitant or crystallization reagent solution—on a siliconized glass cover slide. 0.5 to 1.0 ml of mother liquor is placed in the well of a milliliter plate, and the cover slide is inverted and placed over the well. A seal is formed by a small amount of vacuum grease on the top edges of the well, which also serves to hold the cover slide in place. These clear wells, typically in trays of 24, allow quick visualization of many experiments under a microscope. The mother liquor in the drop is diluted two-fold by the addition of the macromolecule solution, but the drop is small compared to the reservoir. Therefore, in order for this closed system to reach equilibrium, the water from the drop diffuses out to the reservoir over time. This effectively concentrates all of the

components of the drop. If the protein starting concentration is adequately high, it may reach supersaturation levels, which is dependent on the efficacy of the mother liquor.

A great number of experiments are usually required to determine the optimum crystallization conditions for a given macromolecule, and there is no guarantee that any will be found. Finding an appropriate mother liquor is an empirical process, as all proteins have unique chemical composition and structure. There are many variables one can alter to try to achieve supersaturation in a set of hanging drop experiments. The most common are the ionic strength, pH, or temperature of a solution. Another common parameter to change is the identity of the precipitating agent, such as organic solvents or polymers, which, among other effects, decrease the dielectric of the solution.

Ionic strength is a particularly interesting variable because proteins are first "salted in" and then "salted out" with increasing ionic strength (Figure 2.3). Salting out, an effect dependent mainly on anions, can be used to achieve protein supersaturation because salt ions and macromolecules compete for water. Both require hydration layers to maintain solubility, but at some level of salt concentration, proteins must self-associate to satisfy their electronic requirements. In the region of the curves in Figure 2.3 that describe salting-out, the logarithm of protein solubility as a function of ionic strength is approximately linear and can be fitted by the relationship

$$\log S = \beta - K_s (I/2)$$

where  $S$  is solubility,  $K_s$  is the slope of the solubility, relating solubility to ionic strength,  $I$ , and  $\beta$  is the intercept when solubility is extrapolated to zero (Cohn and Ferry, 1943). A similar fit can be obtained for other precipitants, for example, by substituting the concentration of polyethylene glycol (PEG) for  $(I/2)$  and calculating a new  $K$ . The



crystallization of the PurR-hypoxanthine-*purF* operator complex is dependent upon both salting out (0.4 M ammonium sulfate) and addition of polyethylene glycol (PEG) 4000.

PEG 4000 and other large polymers are very successful crystallization reagents, second only in efficacy to ammonium sulfate. PEG molecules are so large that they change the properties of the solution, "crowding out" other macromolecules by volume exclusion. PEGs also lower the dielectric of the medium, thus increasing the electrostatic forces between macromolecules. PEGs work well at low ionic strength, which promotes ligand-binding in macromolecular complexes. The sizes of PEGs used in crystallization range from PEG 100 to PEG 20,000, adding another useful variable to the empirical search for a suitable crystallization reagent.

The goal of all of these empirical trials is a data quality crystal. This means a crystal that is single, has few defects, and is large enough for manipulation and data collection. Crystal volume is proportionately related to diffraction limit, but the intense X-rays generated by third generation synchrotrons can sometimes overcome this restriction. X-ray intensity data can presently be collected on crystals from 0.2 to several mm in length on each side. Once a suitable crystal has been identified, the next goal of the crystallographer is to determine its symmetry.

#### Crystals: symmetry and space groups

Crystals are ordered arrays of molecules that repeat "infinitely" in three dimensions. The ordered repeats can be placed into a volume described by a parallelepiped. The unit cell is defined as having edges defined by three vectors, **a**, **b**, and **c**, with angles between them of  $\alpha$ ,  $\beta$ , and  $\gamma$  (Figure 2.4). The unit cell can be placed

into one of the seven crystal systems. These are characterized by their inherent symmetries, which can dictate the values of each unit cell parameter (Table 2.1). Triclinic is the most general, with the least symmetry; all six lattice parameters can have any value. Certain conventions are followed in the others. For example, in monoclinic crystals, one axis—the principal axis—is perpendicular to the others and contains a 2-fold symmetry axis. By convention, this is the **b** axis. Thus, in monoclinic crystals,  $\alpha = \gamma = 90^\circ$ . Crystals of the corepressor binding domain (CBD) of PurR are monoclinic, and these right angles are found in those crystals, as seen in Table 3.2 of this work. The PurR-hypoxanthine-*purF* operator complex crystallizes in an orthorhombic lattice. In this lattice type, there are three mutually perpendicular 2-fold axes,  $\mathbf{a} \neq \mathbf{b} \neq \mathbf{c}$ , and  $\alpha = \beta = \gamma = 90^\circ$ .

If we assume there is a point in each corner of the unit cell, then we are considering a primitive lattice, denoted P. PurR CBD crystals are one example of this type of lattice. Other, "nonprimitive" types of lattice are possible that will still meet the requirements of the seven crystal systems shown in Figure 2.5. In a cubic (or tetragonal or orthorhombic) system, if a point is placed at the center of the cube, the lattice is termed "body centered" and designated I. If the point is placed at the center of each of the six faces, the lattice is termed F, for "face centered". Monoclinic and orthorhombic lattices can have an additional point at  $1/2, 1/2$  on one of their faces, resulting in a centered face, denoted C by convention. The PurR holo complex crystals have this last type of lattice. When the seven nonprimitive systems are added to the seven primitive ones, the result is the 14 Bravais lattices (Figure 2.5).

Beyond the Bravais lattices, there are 32 classes of symmetry (point groups) that are generated by three basic types of symmetry operators: i) mirror plane, ii) rotation axis, and iii) inversion about a point. Two translational symmetry operations are also possible, the glide plane and the screw axis. The rotation axis is defined such that an  $x$ -fold rotation axis rotates an object  $360^\circ/x$  about the axis (Figure 2.6A) (Stout and Jensen, 1989). Rotation axes with values of  $x$  of 1, 2, 3, 4, or 6 are possible in crystals. The screw axis is defined such that the structure is rotated through an angle corresponding to the degree of the rotation axis and translated parallel to the rotation axis, much like the threads on a screw (Figure 2.6B). Combining the 32 point groups (representing external symmetry) with the 14 Bravais lattices (of internal symmetry) results in 230 unique space groups. However, biological molecules are enantiomorphic in that proteins contain only L-amino acids, so rotation and screw axes are the only possible symmetry operations within crystals of biological macromolecules. This reduces the total number of possible space groups for proteins from 230 to 65. Space groups are signified by a capital letter indicating the lattice type and other symbols representing symmetry operations that are carried out within a unit cell.

The asymmetric unit of a space group is the basic repeating object within a unit cell. One asymmetric unit is thus related to all the other asymmetric units in the unit cell by the symmetry operations just described. It is related to the contents of other unit cells by translations of  $a$ ,  $b$ , and  $c$ . For example, in space groups  $C222$  and  $C222_1$ , there will always be eight asymmetric units in the unit cell, rather than the four found in  $P222$  crystals, due to the addition of centered lattice positions. Because they have identical contents, each asymmetric unit can be used to generate the others, meaning that the

crystallographer need only solve the structure of the asymmetric unit and not necessarily the entire unit cell. Each asymmetric unit may contain a portion of a molecule, if that molecule has internal symmetry, or several independent molecules. PurR CBD, discussed in Chapter 3 of this thesis, crystallizes with one complete dimer in the asymmetric unit. However, the PurR-purine-DNA holo complex lies in a "special position". The symmetry axis of the molecular homodimer coincides with a crystallographic 2-fold symmetry axis, so the asymmetric unit contains one PurR monomer bound to one strand of the DNA pseudopalindrome.

#### X-rays: diffraction and the reciprocal lattice

As noted above, the same year that Max von Laue saw the diffraction of X-rays by crystals, W.L. Bragg was able to describe the scattering observed in terms of reflections from planes of atoms in the crystal lattice (Figure 2.7). His equation, now called Bragg's Law, is:

$$2d \sin\theta = n\lambda$$

where  $d$  is the separation between successive planes of the crystal lattice,  $\theta$  is the angle of reflection of the X-rays,  $n$  is an integer, and  $\lambda$  is the wavelength of the rays (Stout and Jensen, 1989). While X-ray wavelengths can range from 0.1 to 100 Å,  $\lambda$  is 1.54 Å for the Cu  $K_{\alpha}$  radiation used in almost all in-house X-ray generators, including those in the Brennan laboratory. Six of the eight crystal structures reported in this work were collected in-house, while data for two of the structures in Chapter 3 (Table 3.2) were collected using  $\lambda = 1.08$  Å at the Stanford Synchrotron Radiation Laboratory. As used today, Bragg's law describes the reflections of a fixed incident beam as they occur from

planes that are at an angle  $\theta$  relative to the beam, thus generating a reflected ray that deviates through  $2\theta$ .

The X-rays that are scattered from electron density that does not lie in a plane P, as in an actual crystal, can be added to give a resultant wave. This wave accounts for the differing intensities of reflections observed for different planes. Thousands of planes make up each of the mosaic blocks that combine to form a macroscopic crystal. Because of this, diffraction maxima will be sharp and will only occur at clearly defined values of  $\theta$ . Scattered X-rays can be recombined analytically, but in order to reconstruct an image from a diffraction pattern, both the intensity and the phase of each diffracted ray must be known. When crystals are illuminated by X-rays, the electromagnetic rays scatter off of electrons, and the scattered rays may be recorded on photographic film or a number of newer devices. Because only the intensities can be measured, the phase information must be obtained by some other method, resulting in the infamous "phase problem".

The diffraction pattern of a single molecule is its molecular transform. It is also the Fourier transform of the molecule. In a crystal, the scattering of any one molecule is reinforced by the scattering of all the others. However, in the context of a lattice, another level of scattering is introduced. The Fourier transform (FT) of the convolution of two functions is the product of their FTs (Blundell and Johnson, 1976). Conversely, the FT of the product of two functions is the convolution of the transform of one with the transform of the other (Figure 2.8). Likewise, the diffraction pattern of a crystal is the product of the diffraction pattern of the molecule (the molecular transform) and the diffraction pattern of the lattice (the reciprocal lattice). Therefore, the diffraction pattern of the

crystal, the product of these two, is a sampling of the molecular transform at each of the reciprocal lattice (R.L.) points.

The R.L. is a construction that is designed to facilitate the interpretation of diffraction patterns. In terms of Bragg's Law, it is based on  $1/d$ , which varies with  $\sin\theta$ . As seen in Figure 2.9, the R.L. is constructed by taking normals to all "real" planes with indices  $(h, k, l)$  that radiate from some lattice point defined as the origin. The normals have length  $1/d_{hkl}$ , where  $d_{hkl}$  is the perpendicular distance between the planes  $(hkl)$ . R.L. parameters are typically marked with an asterisk, as in Figure 2.9. Whereas the direct unit cell has parameters  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ ,  $\alpha$ ,  $\beta$ , and  $\gamma$ , the R.L. has parameters  $\mathbf{a}^*$ ,  $\mathbf{b}^*$ ,  $\mathbf{c}^*$ ,  $\alpha^*$ ,  $\beta^*$ , and  $\gamma^*$ .

In 1921, P.P. Ewald constructed a geometrical representation of diffraction that satisfies Bragg's Law (Figure 2.10A, reviewed in (Ewald, 1969)). The crystal (C) is placed at the center of a sphere of radius  $1/\lambda$ . The R.L. origin is placed at the point O where an unreflected X-ray would meet the sphere travelling in the direction AC. X-rays are diffracted in the direction CB only if the point B is an R.L. point  $(h, k, l)$ . If this is the case, then the vector OB is an R.L. vector, with  $\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$ . In this construction, the vector  $\mathbf{S}$  is equal to  $1/d_{hkl}$ , and  $\sin\theta = \lambda/2d_{hkl}$ , so that  $\lambda = 2d \sin\theta$ , which is Bragg's Law with  $n = 1$ . If the circle shown in Figure 2.9A is rotated about the diameter AO, the "sphere of reflection" is generated, shown in Figure 2.10B. When the lattice is rotated about its origin, different R.L. points become coincident with the sphere of reflection and their corresponding reflections can be observed. According to Friedel's Law,  $I_{hkl} = I_{\bar{h}\bar{k}\bar{l}}$ , meaning that the intensities reflected from opposite sides of the same set of planes are

equivalent. Because  $\sqrt{I} \propto F$ , where  $F$  is the magnitude of the structure factor, the absolute value of  $F$  should also be equivalent.

### The structure factor

The structure factor is the resultant of  $N$  waves scattered in the direction of the reflection  $hkl$  by the  $N$  atoms of the unit cell (Stout and Jensen, 1989). Each wave has an amplitude that is proportional to the scattering factor of the atom,  $f_j$ . Each wave also has a phase angle,  $\delta_j$ , with respect to the origin of the unit cell. The structure factor can be calculated such that the phases are expressed in terms of the positions of the atoms and the indices of the reflection. The total phase difference, in radians, between the origin and a point  $(x, y, z)$  is given by  $\delta = 2\pi (hx + ky + lz)$ , because the phase difference between reflections from successive planes of any given set  $hkl$  is  $2\pi$  radians. This total phase difference is substituted into the general term for the structure factor amplitude, which is the absolute value of the resultant vector:

$$|F_{hkl}| = \left\{ \left[ \sum f_j \cos 2\pi(hx_j + ky_j + lz_j) \right]^2 + \left[ \sum f_j \sin 2\pi(hx_j + ky_j + lz_j) \right]^2 \right\}^{1/2}$$

This can be expressed in exponential form as:

$$F_{hkl} = \sum_N f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

where  $f_j$  is the scattering factor of the  $j^{\text{th}}$  atom,  $hkl$  are the indices of each reflection in reciprocal space, and  $x_j$ ,  $y_j$ , and  $z_j$  are the coordinates of each  $j^{\text{th}}$  atom in the unit cell. It is more useful to express the equation in a form that describes  $F_{hkl}$  in terms of the contribution from each volume element of electron density in the unit cell:

$$F_{hkl} = \int \int \int \rho(x, y, z) e^{2\pi i(hx + ky + lz)} dz dy dx = \int_V \rho(x, y, z) e^{2\pi i(hx + ky + lz)} dV$$

Because  $F_{hkl}$  is a Fourier transform of the electron density, it is reversible, so electron density is calculated by:

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

The electron density represents the structure of the protein as "visualized" by X-ray crystallography.

### Assessing the quality of data collection

There are several statistics used to assess the quality of a particular data set. I report  $R_{\text{sym}}$ ,  $I/\sigma(I)$ , and percent completeness in tables of data collection statistics in this work (Table 3.2, 4.2, 5.2).  $R_{\text{sym}}(I)$  is a useful factor for comparing the intensities of symmetry-related reflections. While Friedel's Law states that these should be equal, variations always occur in data sets from macromolecular crystals. These variations occur for a variety of reasons, including differences in absorption between regions of a crystal or simple detector error.  $R_{\text{sym}}$  is calculated by:

$$R_{\text{sym}}(I) = \frac{\sum_{hkl} \sum_i |I_{ihkl} - \overline{I_{hkl}}|}{\sum_{hkl} \sum_i I_{ihkl}}$$

for  $n$  independent reflections and  $i$  observations of a given reflection.  $\overline{I_{hkl}}$  is the average intensity of the  $i$  observations. The lower the value of  $R_{\text{sym}}(\%)$ , the more accurate the data. Typical and acceptable  $R_{\text{sym}}$  values range from 4 to ~15%. Qualitatively,  $I/\sigma(I)$  is the intensity of a particular reflection or set of reflections over the background. Formally,



$I = N_p - N_b$  and  $\sigma(I) = (N_p + N_b)^{1/2}$ , where  $N_p$  is the mean value of peak intensity observations and  $N_b$  is the mean value of background intensity observations. The completeness of a data set refers to the number of observed unique reflections collected in a given experiment divided by the total number of unique reflections.

### Phase determination

A number of methods for determining phase angles are available to the crystallographer. Usually, the more straightforward ones are attempted first. The method I have employed in determination of the structures reported in Chapters 3-5 is that of molecular modification, by far the most straightforward method. The other methods, which I shall mention only briefly, are treated in detail in several textbooks ((Blundell and Johnson, 1976; Drenth, 1994; Stout and Jensen, 1989), for example).

Among the other methods, multiple isomorphous replacement (MIR) requires no prior knowledge of the structure within the crystal. MIR does require that data sets be collected on one nonderivatized, "native" crystal and on at least two crystals derivatized by different heavy atoms. Relative heavy atom positions are determined through difference Patterson functions, and these parameters are used to calculate the phase angles of the macromolecule. Another *de novo* method, multiple anomalous dispersion (MAD), is becoming increasingly popular due to the availability of synchrotron beamlines and freedom from determining the empirical and approximate, if any, heavy atom soaking conditions. This method relies on the anomalous scattering of particular atoms, such as selenium, at different wavelengths. With MAD, data sets for a protein crystal, grown as a selenomethionine-substituted molecule, are collected at different

wavelengths at a synchrotron, which has the advantage of tunable, i.e., multiple, wavelengths. Data sets are typically collected at the absorption edge ( $K_{se} \lambda = 0.98 \text{ \AA}$ ), at the inflection point, and at some remote wavelength. An anomalous Patterson map is calculated, and the heavy atom positions are determined. Again, these heavy atom parameters are used to determine phases. A third method of phase determination, termed molecular replacement (MR), requires that a structure homologous to the one under study has been previously determined. The method of MR involves searching the unit cell of the unknown structure with all or part of a homologous structure. This can be done by rotational followed by translational searches, as implemented in the computer program MERLOT (Fitzgerald, 1988). Alternatively, a genetic algorithm can be applied, as in the program EPMR, in which the three rotational and three translational parameters are optimized simultaneously (Kissinger et al., 1999). The genetic algorithm mates populations of test parameters, allowing mutation, and tests their fitness over cycles of "evolution". However, using the structure of the exact same protein in a different conformation as a search model does not guarantee a solution.

The method of molecular modification is much more facile than any of those mentioned above. It is used primarily when determining the structure of a mutant that is isomorphous to the wild-type protein. The experimental amplitudes and phases of the homologous structure (the model) are used directly to calculate the structure factors. Difference Fourier electron density maps are calculated as follows:

$$\text{"}2f_o - f_c \text{ map"}: \quad \rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l (2|F_{obs}| - |F_{calc}|) e^{-2\pi i(hx + ky + lz) + i\alpha_{model}}$$

$$\text{"}f_o - f_c \text{ map"}: \quad \rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l (|F_{obs}| - |F_{calc}|) e^{-2\pi i(hx + ky + lz) + i\alpha_{model}}$$

where  $\rho(x,y,z)$  is the electron density term, and  $2f_o - f_c$  and  $f_o - f_c$  are the maps calculated. The crystallographer then inspects the maps, looking for changes in the modified protein. If there are few changes, then refinement of the model is begun.

### Model refinement

Many refinement software packages are available, but until recently, only the TNT suite of programs was used in the Brennan laboratory (Tronrud et al., 1987). I have used TNT to refine each of the structures I solved that are included in this work. TNT uses Fast Fourier Transforms (FFT) to minimize functions using least-squares refinement.

Refinement is initiated with cycles of "rigid body" refinement until convergence is reached. In this phase, a rigid geometry is assigned to all or part of the model structure, and the parameters of these constrained parts are refined rather than the individual atomic parameters. An entire molecule can be regarded as a rigid entity, or it can be broken down into domains, if they are known. In PurR holo complex refinement, it is useful to regard first the protein and DNA separately, then the CBD, DBD, and DNA separately. The quantity minimized is:

$$Q = \sum_{hkl} w_{hkl} \{ |F_{obs}(hkl)| - |F_{calc}(hkl)| \}^2$$

where  $w$  is a weighting factor. This is done with moderate resolution data (3.5 - 4.0 Å and below), as limiting the resolution of the data in rigid body refinement broadens the  $\Sigma w \Delta F^2$  space that corresponds to the correct structure.

As refinement progresses, the higher resolution data are included to obtain the most precise values of the adjustable parameters, such as atomic positions (x, y, z). Least-squares refinement in TNT includes restraints that impose certain known molecular features, treating them as additional observational equations. These terms are weighted differentially and added to the function to be minimized. The equation minimized in TNT is as follows:

$$(A) \quad Q = \sum_{hkl} w_{hkl} \{ |F_{obs}(hkl)| - |F_{calc}(hkl)| \}^2$$

$$(B) \quad + \sum_{\text{dist. } j} w_D(j) (d_j^{\text{ideal}} - d_j^{\text{model}})^2$$

$$(C) \quad + \sum_k \sum_{\substack{\text{planes coplanar} \\ \text{atoms } i}} w_P(i,k) (\mathbf{m}_k \cdot \mathbf{r}_{i,k} - d_k)^2$$

$$(D) \quad + \sum_{\substack{\text{chiral} \\ \text{centers } l}} w_C(l) (V_l^{\text{ideal}} - V_l^{\text{model}})^2$$

$$(E) \quad + \sum_{\substack{\text{nonbonded} \\ \text{contacts } m}} w_N(m) (d_m^{\text{min}} - d_m^{\text{model}})^4$$

$$(F) \quad + \sum_{\substack{\text{torsion} \\ \text{angles } t}} w_T(t) (X_t^{\text{ideal}} - X_t^{\text{model}})^2$$

The term in (A) is the crystallographic restraint seen in rigid body refinement. The other terms all correspond to known chemical and stereochemical features, and the ideal values have been derived from many high resolution structures of small molecules. (B) restrains interatomic distances, including bond lengths, angles, and dihedral angles. (C) drives aromatic rings toward planarity. (D) imposes the correct enantiomer upon chiral centers. (E) restrains van der Waals contacts, using only the repulsive term. (F) restrains side chain torsion angles.

Positional refinement (x, y, z) is generally carried out to 3.0 Å resolution. Beyond 3.0 Å, the thermal factor, B, is refined for each atom. B is related to  $\overline{u^2}$ , the mean-square amplitude of atomic vibration, by the equation  $B = 8\pi^2\overline{u^2}$ . The higher an atomic B factor, the more thermal motion displayed by the atom in the crystal.

After refinement, the crystallographer visually inspects the fit of the crystallographic model into the electron density map using a molecular graphics program, such as O (Jones, 1991). In O, the model can be manipulated and the map can be fitted. After making the desired changes, refinement is again carried out. These iterations between map fitting and refinement proceed until the crystallographer deems the structure done. During this time, stereochemical problem areas are fixed and solvent molecules are added if the resolution is sufficient to "see" them.

#### Assessing the quality of a model

There are several parameters that can be used to assess the correctness of a model structure. The one most commonly quoted is the R factor, given by:

$$R_{factor} = \frac{\sum_{hkl} |F_{obs}| - k|F_{calc}|}{\sum_{hkl} |F_{obs}|}$$

This is a good measure of how well the model fits the experimental diffraction data. However, the R factor can be made artificially low by a lack of restraint on other parameters refined. Therefore, it is useful to report the root mean-squared deviation (rmsd) values for the geometric restraints in conjunction with the R factor. Target

values for those rmsds are: bond lengths 0.02 Å, bond angles 3.00°, trigonal atoms 0.02 Å, planar groups 0.02 Å, and B correlation 6.00 Å<sup>2</sup> (Tronrud et al., 1987).

Another test for the accuracy of the fit of a model into electron density is to inspect "omit maps". These are difference Fourier maps in which part of the atomic model is removed, and the electron density is calculated and refined against the rest of the model. A typical  $F_o - F_c$  electron density difference map is calculated, as described above. The resulting density, which is unbiased by the region of the molecule that was removed, is inspected. The correct position of the deleted region should be apparent in this map.

Ramachandran plots are also good indicators of the stereochemical quality of a model, but they are only applicable to proteins (Laskowski et al., 1993). These are plots of the protein backbone dihedral torsion angles,  $\Phi$  and  $\Psi$ , for each residue. Together,  $\Phi$  and  $\Psi$  can assume a certain range of energetically favorable angles due to steric hindrance. Different types of secondary structure often fall into characteristic locales on the Ramachandran plot, and glycines can occupy a wider range of angles (Figure 2.11A). Good models will have few or no residues outside the most favorable regions of the plot. However, sometimes deviant  $\Phi$  and  $\Psi$  angles are not mistakes in the model but are instead found in strained regions of the backbone (Schumacher et al., 1994a). PurR has two such residues that are found in the wild-type and all mutant structures solved: Ser124 and Asp275 (Figure 2.11B).

Table 2.1. The seven crystal systems. Adapted from (Blundell and Johnson, 1976).

Name	Possible Bravais Lattices	Axes of Symmetry	Lattice	
Triclinic	P	None	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$
Monoclinic	P, C	1 dyad axis (parallel to $b$ )	$a \neq b \neq c$	$\alpha = \gamma = 90^\circ \neq \beta$
Orthorhombic	P, C, I, F	3 dyad axes mutually orthogonal	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Tetragonal	P, I	1 tetrad axis (parallel to $c$ )	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Trigonal	P	1 triad axis (parallel to $c$ )	$a = b \neq c$	$\alpha = \beta = \gamma = 120^\circ$
	(or R)		$a = b = c$	$\alpha = \beta = \gamma < 120^\circ \neq 90^\circ$
Hexagonal	P	1 hexad axis (parallel to $c$ )	$a = b \neq c$	$\alpha = \beta = 120^\circ, \gamma = 90^\circ$
Cubic	P, I, F	4 triad axes (along the diagonals of the cube)	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$

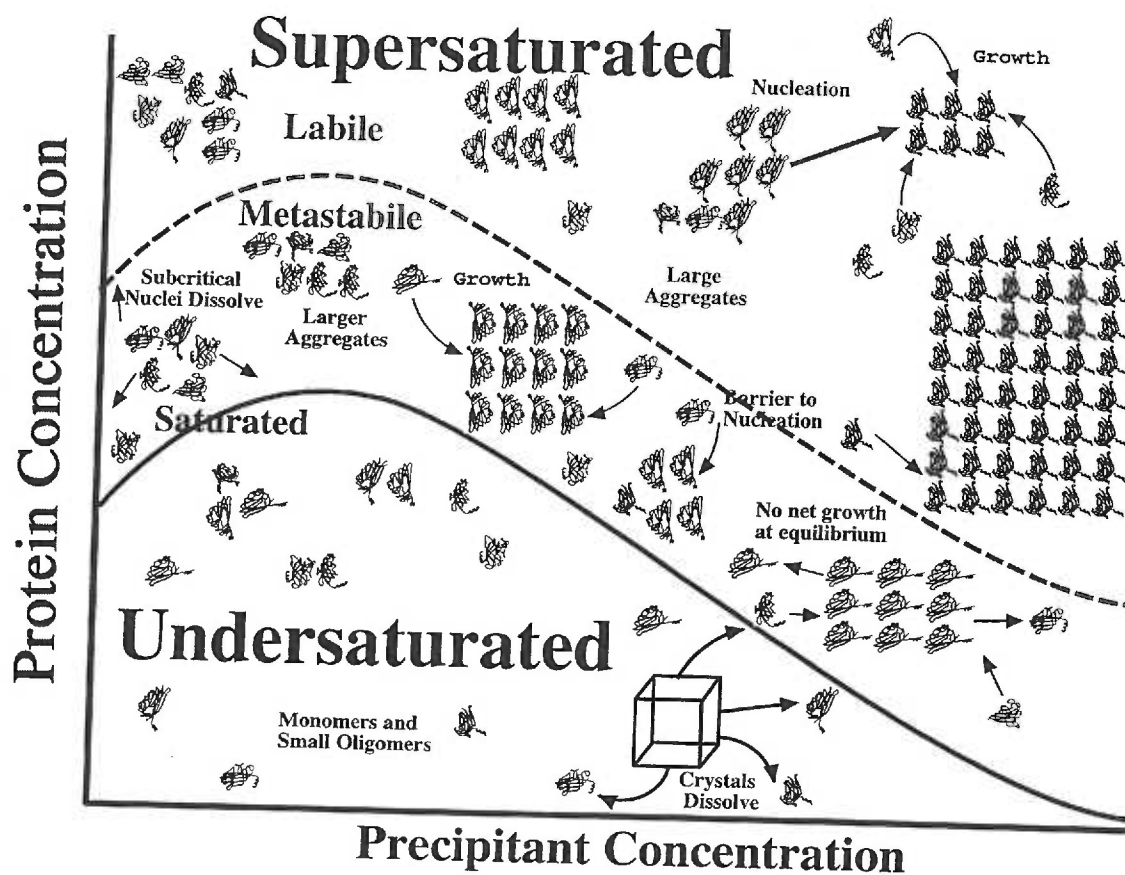


Figure 2.1. The phase diagram for crystallization, depicting the physical events occurring in each region that pertain to crystallization. Adapted from (McPherson, 1999).



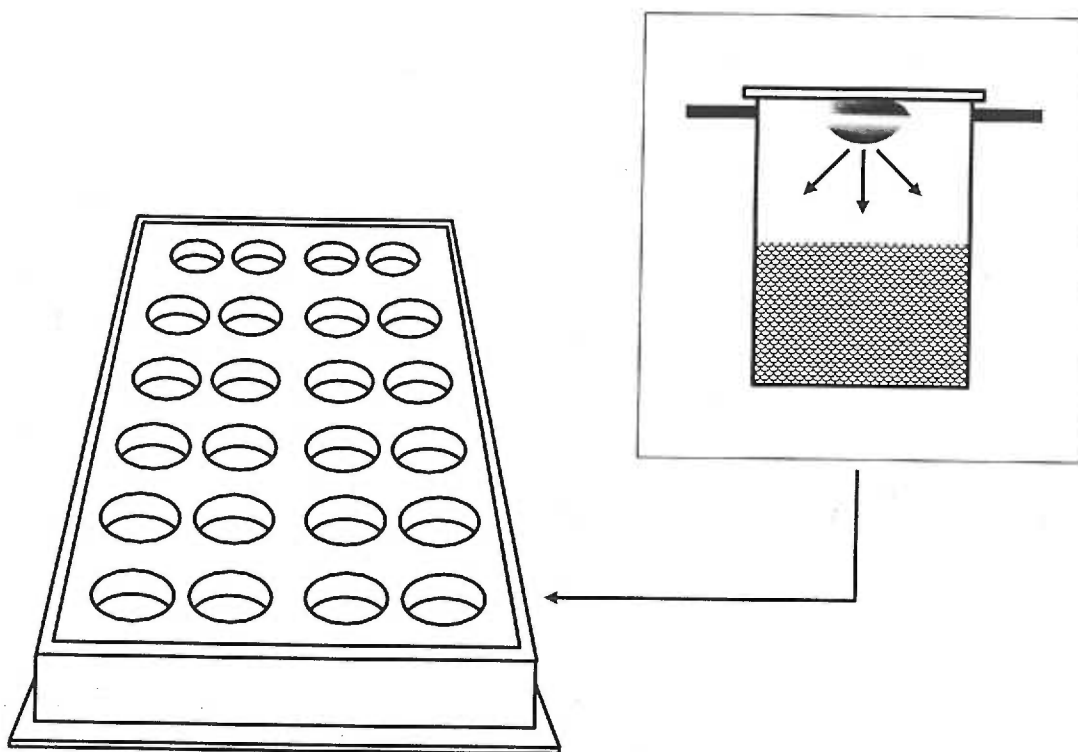


Figure 2.2. A milliliter plate used in hanging drop vapor diffusion experiments. Inset: One well in such an experiment, with the drop hanging from a cover slide, over the mother liquor solution. Adapted from (McPherson, 1999).

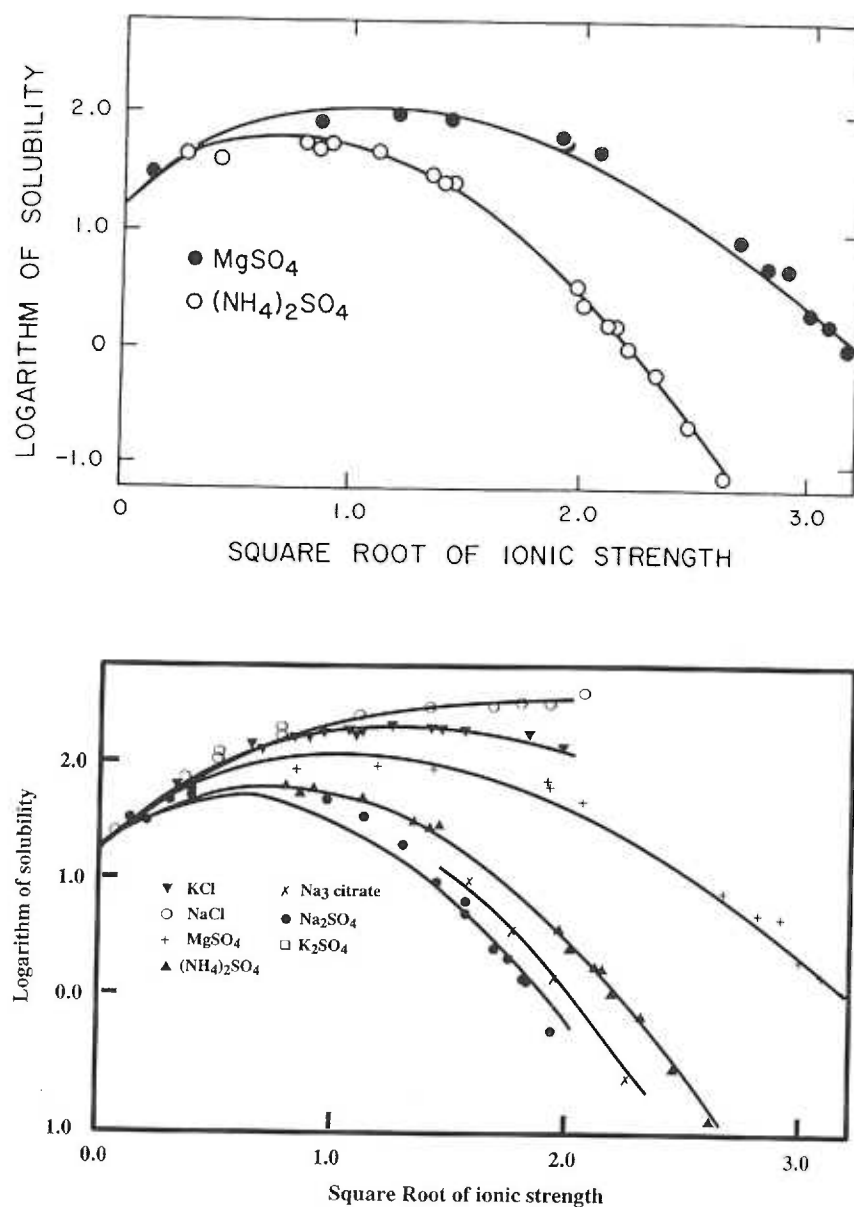


Figure 2.3. Top: The solubility of enolase as a function of ionic strength of two salts. From (McPherson, 1999). Bottom: The solubility of carbonmonoxyhemoglobin in several salt solutions at 25°C. Adapted from (Green, 1932).

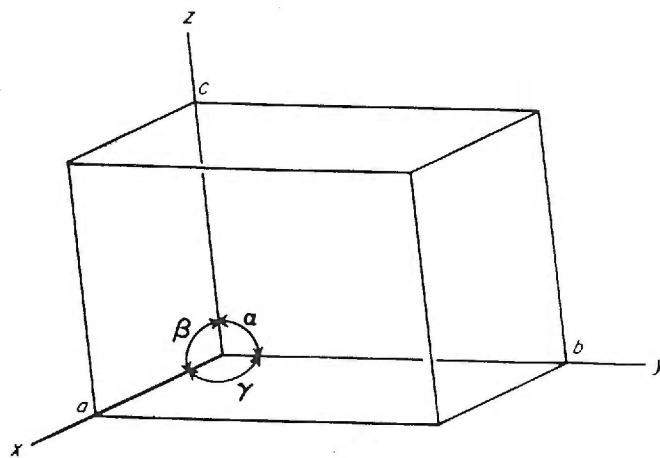


Figure 2.4. The unit cell. From (Stout and Jensen, 1989).

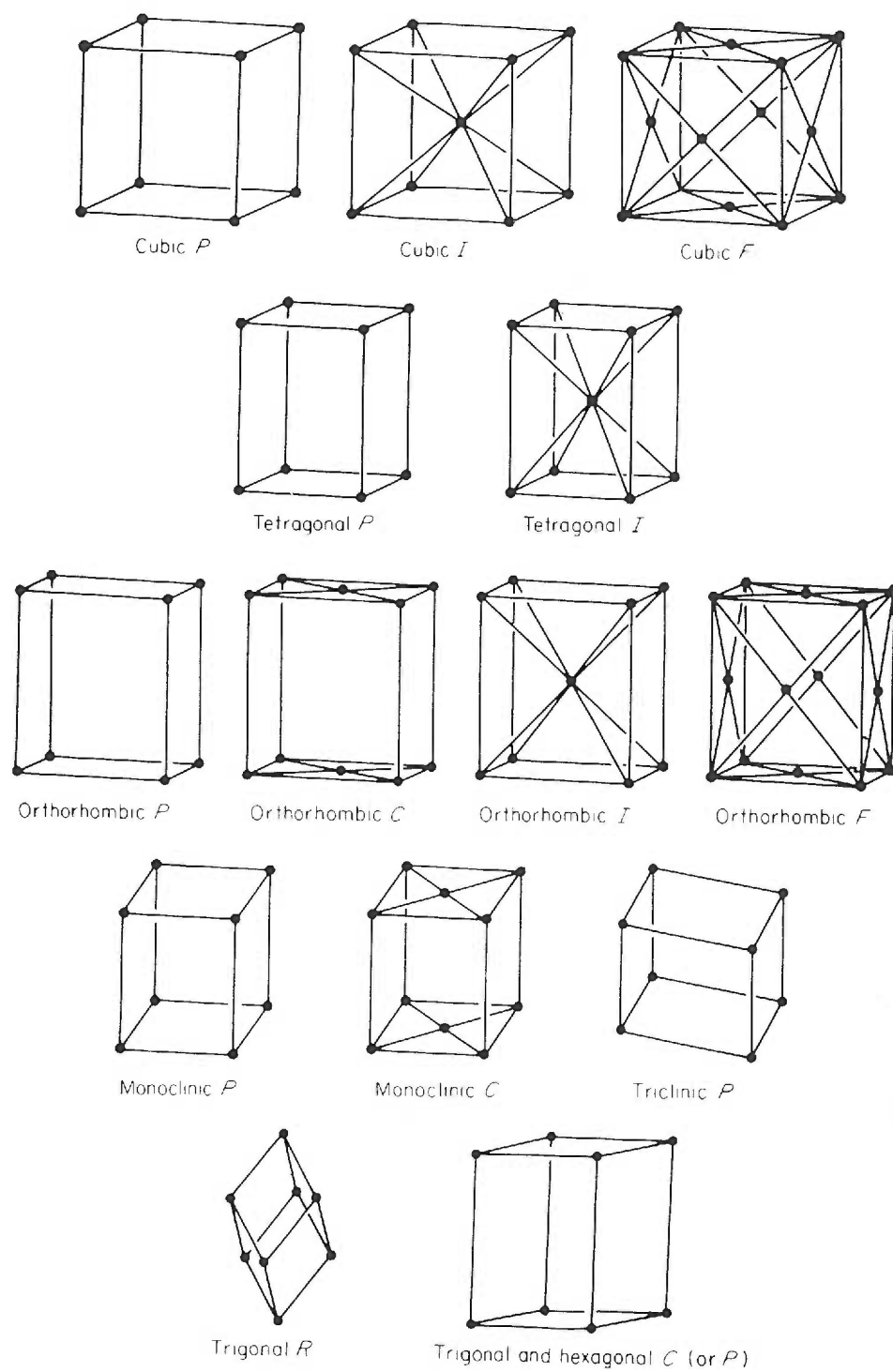


Figure 2.5. The fourteen Bravais lattices. From (Blundell and Johnson, 1976).

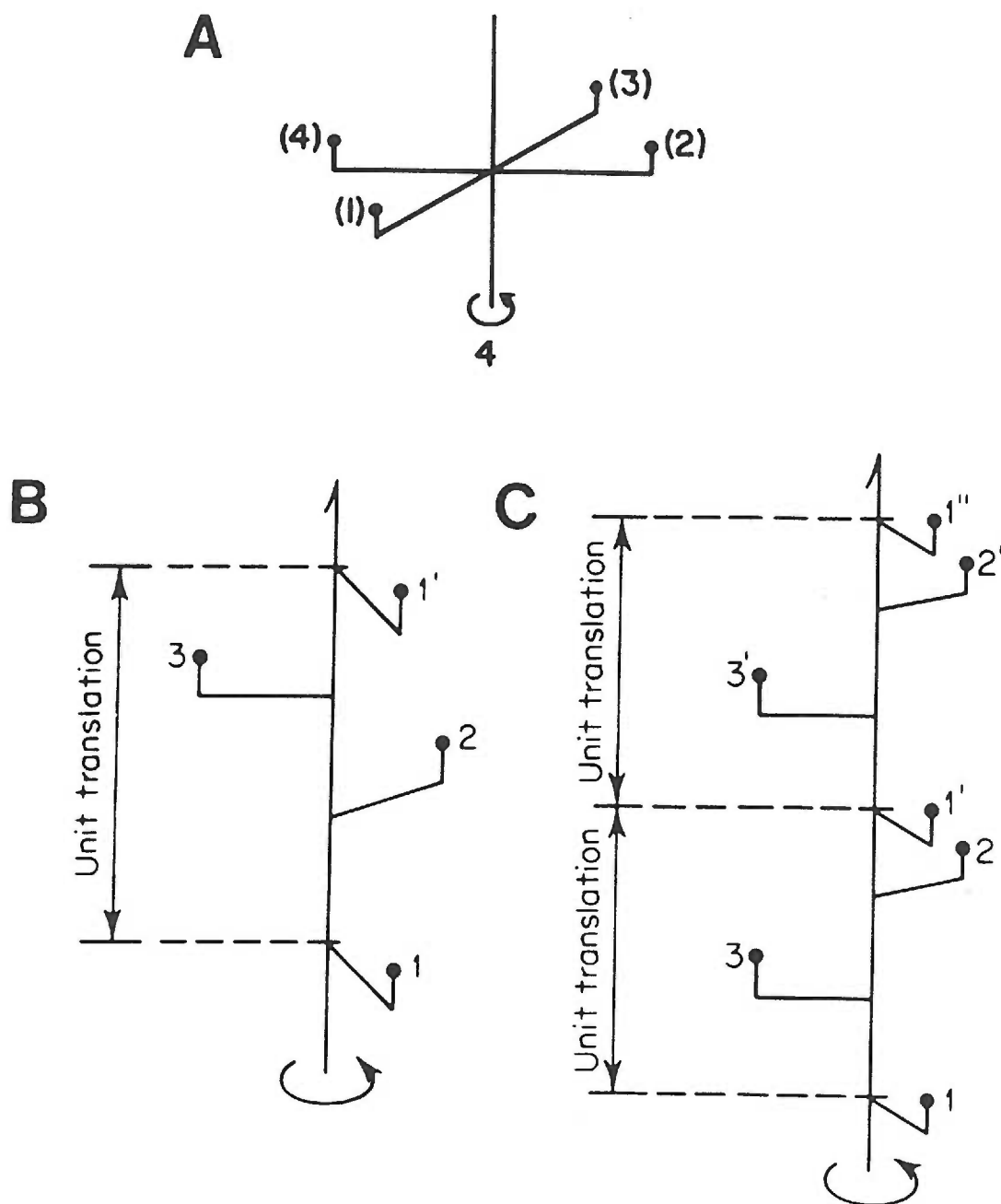


Figure 2.6A. A four-fold rotation axis. B. A  $3_1$  screw axis. C. A  $3_2$  screw axis. These figures were adapted from (Stout and Jensen, 1989).

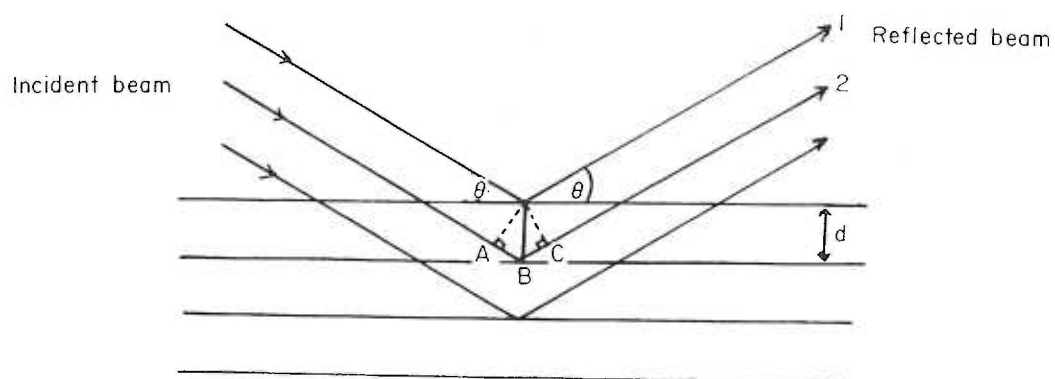


Figure 2.7. An illustration of Bragg's law. From (Blundell and Johnson, 1976).

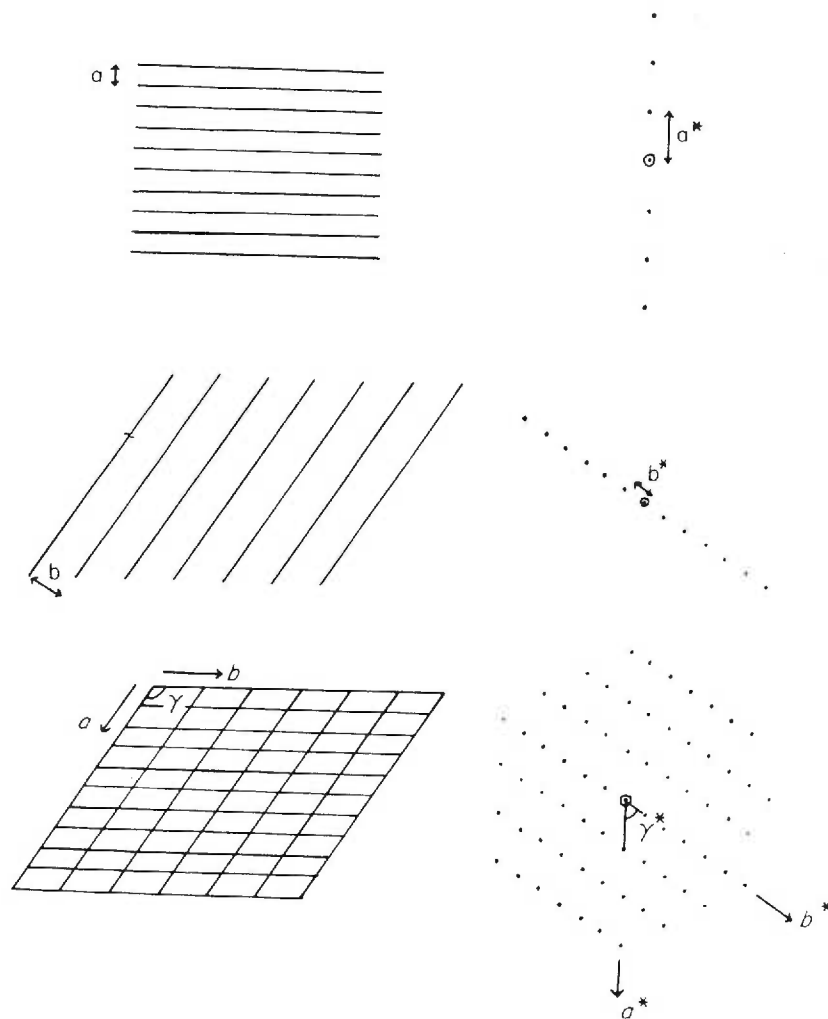


Figure 2.8. The Fourier transforms of a set of lines and a lattice. Note that the transforms of the lines is a row of dots, however the transform of the convolution of the lines is the product of their individual transforms. Adapted from (Holmes and Blow, 1966).

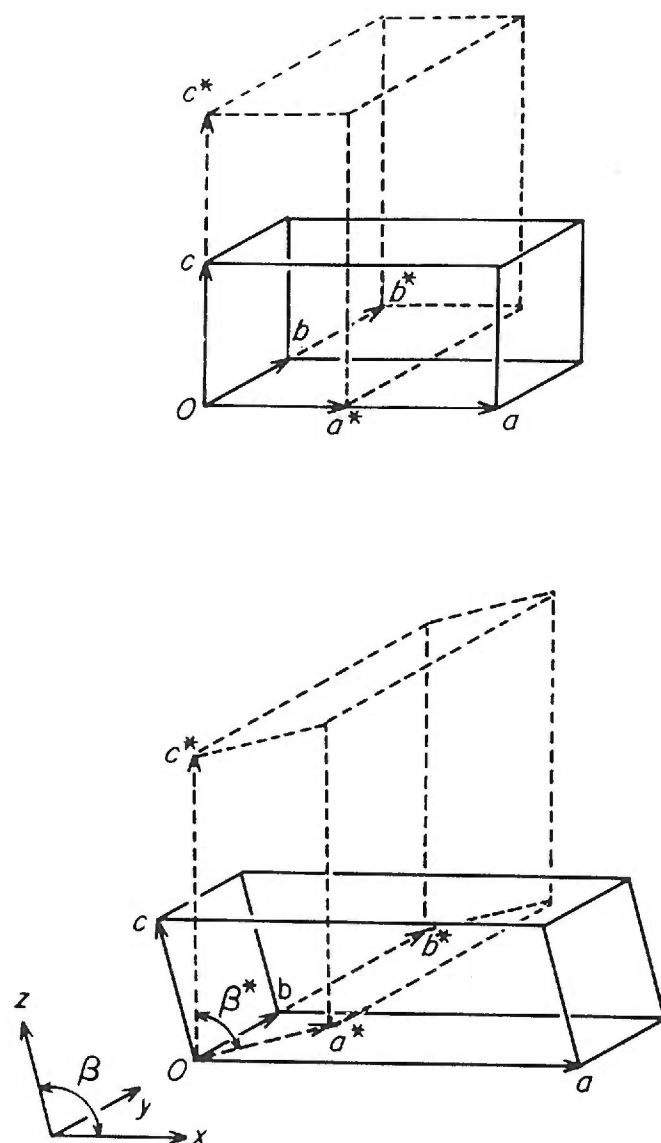


Figure 2.9. Top: Orthorhombic direct (solid lines) and reciprocal cells (dashed lines).

Bottom: Monoclinic direct and reciprocal cells, using the same convention. These figures were adapted from (Stout and Jensen, 1989).



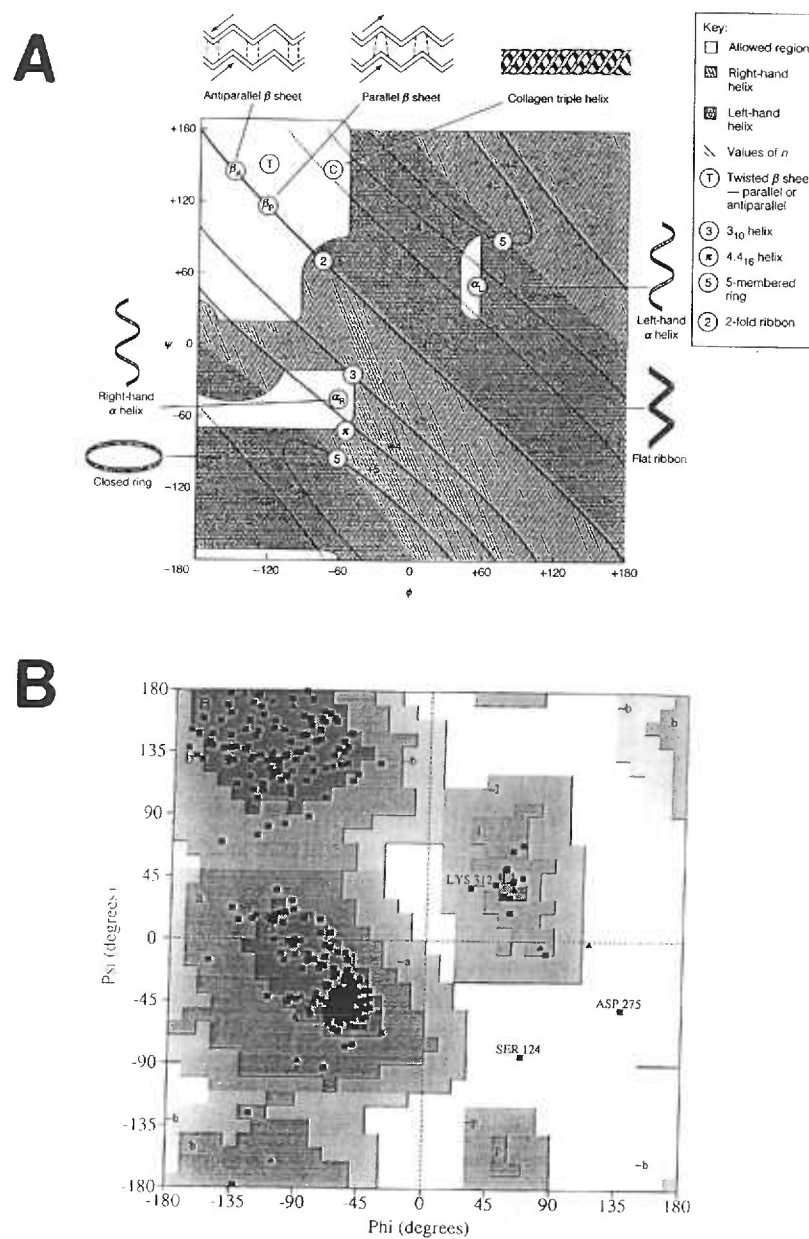


Figure 2.11A. A Ramachandran plot showing allowed regions and regions characteristic of particular types of secondary structure. B. The Ramachandran plot for wild-type PurR bound to hypoxanthine and the *purF* operator.

**Chapter 3. Structural and functional characterization of PurR allosteric mutants  
W147F, W147A, and W147R**

Joy L. Huffman<sup>1</sup>, Howard Zalkin<sup>2</sup>, and Richard G. Brennan<sup>1§</sup>

<sup>1</sup>*Dept. of Biochemistry and Molecular Biology, Oregon Health Sciences University*

<sup>2</sup>*Department of Biology, Purdue University*

## Abstract

The crystal structures of corepressor-bound and free *E. coli* purine repressor (PurR) have delineated the roles of several residues in corepressor binding and specificity and the intramolecular signal transduction (allosterism) of this LacI/GalR family member. From these structures, residue W147 was implicated as key for the allosteric response, but in most members of the LacI/GalR family, position 147 is an arginine. To understand the role of this tryptophan at position 147, three substitutions, to phenylalanine (W147F), alanine (W147A), or arginine (W147R), were constructed and characterized *in vivo* and *in vitro*, and their structures were determined. W147F displays a decreased affinity for corepressor and is a poor repressor *in vivo*. W147A and W147R, on the other hand, are super repressors, and bind corepressor at least five times more tightly than wild-type. Each of the mutant PurR-hypoxanthine-*purF* operator holo complexes crystallizes isomorphously to wild-type. Interestingly, only the apo corepressor binding domain (CBD) of W147F crystallizes under those conditions used for the wild-type protein. Although screened extensively, neither the apo CBD of W147R nor W147A crystallizes. Structures of the holo repressor mutants have been solved to between 2.5 Å and 2.9 Å, and the structure of the apo CBD of W147F has been solved to 2.4 Å resolution. These structures provide further insight into the mechanism by which W147 transduces the allosteric signal within PurR and a likely explanation for the need of a tryptophan.

## Introduction

The *E. coli* Purine repressor, PurR, is a member of the LacI/GalR family of transcription repressors (reviewed in (Zalkin and Nygaard, 1996)). PurR is a 341 amino acid residue, 38 kDa protein that is always found as a homodimer. PurR consists of two domains, an N-terminal DNA-binding domain (DBD) encompassing residues 1-60 and a C-terminal corepressor binding domain (CBD) encompassing residues 61-341. The CBD is composed of two subdomains, N-CBD and C-CBD. PurR regulates the transcription of at least 18 genes that encode enzymes involved primarily in the *de novo* biosynthesis of purine nucleotides. A corepressor, hypoxanthine or guanine, must first bind to PurR to induce a conformational change, which enables PurR to bind its 16 base pair operator site, thereby effecting transcription repression. Thus, the environmental signal of excess purine is transduced into a biological response, namely repression of *de novo* purine nucleotide biosynthesis.

Several crystal structures of PurR-corepressor-DNA ternary complexes and the corepressor-free form of the CBD have been solved previous to this work (Glasfeld et al., 1996; Lu et al., 1998b; Schumacher et al., 1995; Schumacher et al., 1994a). These crystal structures reveal residues involved in DNA or corepressor binding and support a route by which the intramolecular signal of corepressor binding is passed from the corepressor binding domain to the DNA binding domain, which is over 40 Å from the corepressor binding pocket. Upon binding corepressor, structural changes that include rotations of at least 17° between the NH<sub>2</sub> and COOH-subdomains of the CBD of each monomer result in its "closed" conformation (Figure 3.1A, B). These rotations restructure the interface of the N-subdomains of the CBD, which in turn positions the DNA-binding domains of each

PurR monomer such that in the presence of operator DNA, the hinge region, residues 48-56, can undergo a coil to helix transition and DNA can subsequently be bound.

The structures of corepressor-bound and free PurR have revealed residue W147 as one of the keys to intramolecular signal transduction, i.e. the allosteric response, yet the bound purine corepressor does not interact with W147. The critical role of this residue is in the corepressor free form of the protein, in which W147 is found in the corepressor binding pocket, where it stacks with Y73 and F74 and acts as a structural surrogate for corepressor (Figure 3.1D). The W147 Nε also makes a hydrogen bond with the hydroxyl group of Y73, further stabilizing its position in the pocket. Upon binding corepressor, W147 moves up out of the pocket to stack with Y126. This translocation is over 9 Å (for W147 Cζ3 and also alters the position of residues 146-159 (Figure 3.1C). It is this motion that leads to the change in the monomer-monomer interface of the N-subdomain.

To dissect the role of the indole side chain of PurR residue 147, three substitutions (to F, A, or R) were constructed, their biological and biochemical effects were characterized *in vitro* and *in vivo*, and their germane structures were determined. Additionally, we modelled the CBD of W147A and W147R in their corepressor free forms, because they could not be crystallized. The phenylalanine replacement was designed to test the effect of removing the ability to hydrogen bond while retaining the hydrophobicity and ability to make protein-style aromatic stacking interactions found with the wild-type tryptophan. The W147A mutant was made to examine the effect of removing the indole side chain altogether. The arginine replacement at position 147 was designed to conserve the length and ability to hydrogen bond while removing the stacking capability and hydrophobicity of the wild-type tryptophan.

The *in vivo* repression results revealed an interesting behavior: W147F was a weaker repressor than wild-type, while W147A and W147R were both "super repressors". The *in vivo* differences correlated well with *in vitro* binding studies, which demonstrated that altered affinities for corepressor were largely responsible for these changes in repression. The structures reveal changes localized about residue 147 that account for the differences in corepressor affinity. The combination of the binding assays and crystal structures of W147 mutants provides insight into the allosteric regulation of PurR and more generally of DNA binding proteins.

## Materials and Methods

### *Assay for repression by PurR.*

FL100 *purR* integrants were cultured overnight at 37° C in minimal medium (Rolfes and Zalkin, 1990b) containing 15 µg/ml chloramphenicol, which was diluted 1:50 to inoculate the same medium. This culture was grown, shaking at 37° C until mid-log phase was reached. The β-galactosidase activity was assayed in permeabilized cells according to the method of Miller (Miller, 1972).

### *Overexpression and purification of PurR.*

All mutant protein overexpression plasmids were generously provided by the Zalkin laboratory. Overexpressed PurR was purified as described (Choi and Zalkin, 1992). For purified PurR, the protein concentration was determined spectrophotometrically at 280 nm using  $A_{280}$  (1%) of 12.9 (Rolfes and Zalkin, 1990b). An additional purification step was included for the W147A protein used in binding studies, as it had retained some purine base through the normal purification steps. W147A mutant protein was treated with 1 M guanidinium hydrochloride to allow the release of purine base, as described (Lu et al., 1998a).

### *Isolation of the PurR corepressor binding domain.*

The CBD of each mutant was prepared by utilizing a trypsin cleavage site between residues Arg52 and Ser53 in the hinge region of PurR, which lies between the DNA binding domain (DBD) and corepressor binding domain. Trypsin digestion of PurR mutants was carried out as described previously (Choi and Zalkin, 1992).

Modifications were made to the procedure for isolating the CBD from the DBD and trypsin. Instead of adding 5 mM PMSF and passing the solution over a sizing column, the trypsin was removed by a 15 minute incubation of the reaction mixture, shaking and on ice, with trypsin inhibitor crosslinked to agarose beads (Sigma). The beads were pelleted by centrifugation and the supernatant removed. This incubation, followed by removal of beads, was repeated to ensure complete removal of trypsin. The remaining 64 kDa CBD dimer was concentrated to ~10 mg/ml, and its buffer was exchanged to 100 mM NaCl, 5% glycerol, and 20 mM Tris, pH 7.6, using Millipore 50 kDa Molecular Weight cut-off filter devices. This step concentrated the protein for use in crystallization and effectively removed the 6 kDa DBD by filtration. Complete trypsin digestion and removal of the DBD were confirmed by SDS-PAGE.

*PurR equilibrium binding to corepressor and operator DNA.*

Fluorescence polarization experiments were done with a PanVera Beacon Fluorescence Polarization System (PanVera Corporation). Samples were excited at 490 nm and emission was measured at 530 nm. A 5' fluoresceinated oligonucleotide corresponding to the *E. coli purF* operator (Oligos, Etc.) (5'-F-AAAGAAAACGTTTGCGTACCCCCTACGCAAACGTTTTCTTT-3') was self-annealed by heating to 80 °C, followed by flash cooling, to form a stem-loop structure with the *purF* operator motif at the center of the base-paired region. Binding was assayed in a 1 ml volume at 25 °C. The components of each binding experiment were: 2 nM fluoresceinated DNA, 250 mM potassium glutamate, 150 mM NaCl, 10 mM magnesium acetate, 1 mM EDTA, 100 mM potassium HEPES, pH 7.5, and 1.0 µg/ml poly d(IC).



For experiments measuring the binding of protein to *purF* operator, saturating amounts (200  $\mu$ M) of hypoxanthine were included in the cuvette . The PurR protein was diluted serially between 100  $\mu$ M and 0.3  $\mu$ M and titrated into the cuvette. Because PurR does not bind DNA appreciably in the absence of corepressor, the same type of assay was used to measure the binding of corepressor to protein indirectly. In these experiments, saturating amounts of protein (100 nM) were included in the cuvette, and hypoxanthine was titrated. It is assumed that corepressor-bound protein is the only molecular species that can bind the fluoresceinated *purF* operator. Since such binding is immediate, the change in polarization can be measured as a function of indirect corepressor binding. After each addition of either protein or corepressor, samples were incubated in the Beacon instrument at 25 °C for 30 s before a measurement was taken. A 30 s incubation allowed the binding equilibrium to be reached. The observed mP (milliPolarization) at each titration point represents the average of eight measurements integrated over 6 s.

The data of each binding isotherm were analyzed by least squares regression using SigmaPlot software (Jandel Corporation). The following equation was applied:

$$A = \left\{ \frac{(A_{bound} - A_{free})[P]}{K_d + [P]} \right\} + A_{free}$$

where A is the polarization measured at a given total concentration of peptide [P],  $A_{free}$  is the initial polarization of the free DNA, and  $A_{bound}$  is the polarization of maximally bound DNA (Glasfeld et al., 1996; Lundblad et al., 1996). Least squares regression analysis was used to determine  $A_{free}$ ,  $A_{bound}$ , and  $K_d$ .

*Crystallization and data collection.*

PurR mutants are stored at a concentration of 8-10 mg/mL in 200 mM potassium phosphate, pH 7.5, 0.1 mM EDTA, 0.2 mM DTT, 5% glycerol. The 16 bp *purF* operator (Oligos, Etc.) has the sequence 5' - TACGCAAACGTTTTCTT - 3', containing a 5' overhanging T when annealed (Schumacher et al., 1994b), and is stored in 10 mM sodium cacodylate, pH 6.5. Crystals of the PurR mutant-purine base-*purF* operator complex were grown using the hanging drop vapor diffusion method, as described previously (Schumacher et al., 1994b). Briefly, 100  $\mu$ M protein (8 mg/mL) solution saturated with purine base is mixed with *purF* operator so that a 1:2 molar ratio of protein:duplex DNA solution resulted. This solution is then mixed 1:1 with the crystallization solution, which is 25% PEG 4000 (BDH), 0.4 M ammonium sulfate, 50 mM cobalt hexammine chloride, 0.1 M ammonium phosphate, pH 7.5. The crystallization experiments are carried out at 23 °C. Crystals grow first as thin plates at the interface of the phase separation that occurs in the hanging drop. After several weeks to months, these plates yield large (0.6 mm x 0.3 mm x 0.2 mm) rhombohedrons. The resulting crystals are isomorphous to wild-type PurR-hypoxanthine-*purF* operator crystals. The W147R-hypoxanthine-*purF* operator crystal was cryocooled by flash-freezing in liquid nitrogen after a 30 s incubation in a cryoprotectant mother liquor consisting of 25% PEG 4000, 0.4 M ammonium sulfate, 0.1 M ammonium phosphate, pH 7.5, and 35% ethylene glycol.

Crystals of W147F-CBD were grown using hanging drop vapor diffusion. A solution of ~10 mg/ml protein was mixed 1:1 with a mother liquor of 20% PEG 600, 0.2 M MgCl<sub>2</sub>, 0.1 M Tris HCl, pH 7.6. Crystals typically appear after one day but continue to grow for about one week. These crystals are long rods that have dimensions of up to

1.0 mm x 0.3 mm x 0.3 mm. W147F-CBD crystals are isomorphous to the CBD crystals of the wild-type protein.

X-ray intensity data were collected at 22 °C on an ADSC area detector using a Rigaku RU-200 X-ray generator at 40 kV, 150 mA, or at -180 °C on beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) (see Table 6.2).

#### *Structure determination and refinement.*

Structures were solved using difference Fourier techniques utilizing the wild-type PurR-hypoxanthine-*purF* operator structure (IPNR) as the initial model for each of the holo complexes, after removing the hypoxanthine, side chain at position W147, and all solvent. The asymmetric unit contains one PurR monomer, one purine base, and a *purF* operator half-site. The wild-type CBD structure was used as the initial model for the W147F-CBD data.

Refinement of each complex was initiated with rigid body refinement until convergence, followed by XYZ (positional) refinement, and finally XYZ and B (thermal) refinement, as implemented in TNT (Tronrud et al., 1987). Hypoxanthine, the W147X side chains, and solvent molecules were built into the structures in the final stages of refinement using O (Jones, 1991).  $F_o - F_c$  omit maps were calculated for each structure to ensure the unbiased placement of side chains, purines, and solvent molecules. The components in question were removed from the model, which was then refined for 20 cycles or more before inspection of the omit density map. The stereochemical quality of each structure was ascertained using PROCHECK (Laskowski et al., 1993).

Models of the W147A and W147R corepressor binding domains were obtained by simply mutating the tryptophan at position 147 in the wild-type structure to either an alanine or an arginine using the program O (Jones, 1991). To allow reasonable hydrogen bonding and stereochemistry, and to avoid steric clash, the R147 side chain was rotated, but the remainder of the protein was not altered in either mutant model structure.

## Results

### *In vivo function of PurR mutants.*

To further investigate the role of W147, which had been implicated previously in *in vivo* and structural studies (Choi et al., 1994; Lu et al., 1998a; Schumacher et al., 1995) in the allosteric mechanism of PurR, we mutagenized this residue to phenylalanine (W147F), alanine (W147A), or arginine (W147R) and characterized their properties. The W147F mutation conserved the aromaticity of the side chain without allowing hydrogen bond formation with Y73. The W147A mutation was designed to reveal the thermodynamic and structural effects of removing the indole ring. The W147R mutation is conservative with respect to length and ability to hydrogen bond, but adds a positive charge and likely removes the contribution of stacking. Interestingly, arginine is the most common residue found at position 147 in the LacI/GalR family (Markiewicz et al., 1994).

Mutant genes were constructed and integrated into the phage  $\lambda$  attachment site of the *E. coli* chromosome in single copy in order to evaluate repressor function. The *purR* gene in this construction is controlled by a weak synthetic promoter chosen to give expression comparable to the wild-type gene at its normal site. Western blots of wild-type cells and the integrants verified that all strains contained similar PurR levels (not shown). Repression of the *purF* operator-*lacZ* reporter by PurR<sup>+</sup>, W147F, W147A, and W147R was 13.9, 4.9, 33.6, and 31.3-fold, respectively (Choi et al., 1994; Lu et al., 1998a), Table 3.1). W147F displays weaker repression (4.9 fold) than wild-type PurR (13.9 fold). W147A and W147R behave as “super repressors”, displaying much higher *in vivo* repression (31.1 and 33.6 fold) than wild-type PurR.

### *W147F, A, and R equilibrium binding studies.*

To determine the effects of these mutations on the ability of PurR to bind corepressor, we used fluorescence polarization to determine the *in vitro* equilibrium dissociation constants ( $K_d$ s) of each of the W147 mutants and wild-type PurR for both *purF* operator DNA and one of its physiological corepressors, hypoxanthine. Previous studies have employed equilibrium dialysis and electrophoretic mobility shift assays for the determination of corepressor binding affinities (Choi and Zalkin, 1992; Rolfes and Zalkin, 1990b). Our results for wild-type PurR agree well with these reported values [previous results:  $K_d(\text{DNA}) = 0.9 - 2.5 \text{ nM}$ ,  $K_d(\text{hyp}) = 7 - 11 \text{ }\mu\text{M}$ ]. Small variations can be attributed to the use of different methods. Binding data from the present experiments are shown in Figure 3.2, and the data are summarized in Table 3.1.

W147F shows a significant decrease in affinity for corepressor ( $K_d = 24.0 \text{ }\mu\text{M}$ ) and acts as a much weaker repressor than wild-type (4.9 fold, *c.f.* 13.9 for wild-type). In contrast, the W147A and W147R mutants have increased affinities for corepressor (less than  $0.05 \text{ }\mu\text{M}$  and  $0.85 \text{ }\mu\text{M}$ , respectively). Changes in *in vitro* corepressor affinity are well correlated with *in vivo* repression activity and are not a result of an altered DNA affinity. With the exception of W147A, the W147 mutant and wild-type affinities for *purF* operator are essentially identical (2.0 to 2.5 nM). W147A bound *purF* operator with a slightly higher affinity ( $K_d = 0.7 \text{ nM}$ ) than wild-type PurR.

### *Crystallographic studies.*

Mutants W147F, W147A, and W147R were cocrystallized with the *purF* operator and hypoxanthine by standard methods (Schumacher et al., 1994b), and each of these

crystallized isomorphously to wild-type. We solved these structures by molecular modification techniques. Selected crystallographic and refinement statistics are listed in Table 2. Only the W147F mutant crystallized in the apo conformation, i.e., the CBD form (Schumacher et al., 1992). These crystals were isomorphous to wild-type. The statistics for that structure determination are also listed in Table 3.2. *De novo* crystallization attempts with the W147A-CBD and W147R-CBD have been unsuccessful.

*Crystal structures of PurR W147F-hypoxanthine-purF operator and W147F corepressor binding domain.*

To elucidate the stereochemical and structural basis by which W147F binds the natural corepressor, hypoxanthine, with a 15-fold lower affinity than wild-type, we determined the crystal structures of the W147F-hypoxanthine-*purF* operator complex and the apo W147F corepressor binding domain (Figure 3A, B). This position is especially interesting because the W147 residue does not contact the corepressor directly.

As anticipated, the corepressor binding pocket of the W147F-corepressor-*purF* operator complex reveals that none of the wild-type protein-purine interactions are missing, nor does the hypoxanthine take a radically different binding position (data not shown). Superimposition of the C $\alpha$ s of the W147F structure onto the corresponding atoms of the wild type PurR-hypoxanthine-*purF* operator structure (1QPZ) confirms their structural identity and yields a root mean-squared deviation (rmsd) of 0.24 Å (within the error of the coordinates). As observed in the wild type PurR-hypoxanthine-*purF* operator structure, the DNA is kinked about the central CpG base pair step. The DNA bending angle, which was calculated with CURVES (Lavery and Sklenar, 1988) and excluded the

first and last base pairs of the 16 base pair operator because of their poor electron density, is  $52^\circ$ , indicating that the DNA structure is essentially identical to that seen in the PurR-hypoxanthine-*purF* operator structure (Schumacher et al., 1994a). All other DNA properties, e.g., roll and twist angles and rise per nucleotide, were also nearly identical to those found in the wild type protein-corepressor-DNA structure (data not shown).

As expected, the F147 side chain stacks in an orientation similar to that of the W147 indole ring (Figure 3.3A). However, the loop that contains residues 146-149, and thus F147, moves toward Y126, and the Y126 ring is rotated  $25^\circ$  relative to the wild-type structure. The rmsd between wild-type and W147F C $\alpha$  atoms within this loop, residues 146-154, is 0.77 Å, three times that of the whole protein, with the F147 C $\alpha$  translocating 0.89 Å and the G148 C $\alpha$  moving 1.83 Å relative to their wild-type positions. These small changes serve to strengthen the  $\pi$ - $\pi$  protein-style stacking interactions between F147 and Y126. In the W147F structure, K151 is in an orientation that allows similar dispersive contacts to those seen in the wild-type structure, largely between the phenyl ring and K151 C $\beta$  and C $\zeta$ . In the wild-type repressor, favorable van der Waals interactions are made between the indole ring and the C $\beta$  and C $\epsilon$  methylene carbons of K151, whilst its N $\epsilon$  points away from the W147 indole ring. N $\epsilon$  makes one hydrogen bond to Glu149 O $\epsilon$ 2 in both the W147F and wild-type structures.

The structure of the corepressor-free W147F-CBD is globally identical to wild-type, with a calculated rmsd of 0.47 Å between residues 61-341. However, several local changes are apparent that could contribute to an increased stabilization of this conformer (Figure 3.3B). The phenylalanine side chain is located in the same general position as the wild-type tryptophan and stacks with Y73 and F74. To maintain this arrangement, the



Y73 and F74 appear to move closer to the F147 than observed for W147. Additionally, the tyrosine at position 73 is rotated approximately 25° to engage in better  $\pi$ - $\pi$  stacking interactions with the phenyl ring of F147 (Burley and Petsko, 1985). Moreover, the electron density for the Y73 side chain is weak and suggests two possible orientations for the side chain, with one predominant over the other (as depicted in 3.3B). We have built this residue into the strongest density only because the electron density and resolution of the data are insufficient for refinement of high and low occupancy sites. Perhaps the largest structural perturbation is in the isomerization of the peptide linkage between residues 147 and 148. In this mutant, the protein amide linkage between residues 147 and 148 is *cis* instead of *trans* (Figure 33.B). The *cis* conformation allows W147F to form a good hydrogen bond between D146 O $\delta$ 1 and G148 N (2.7 Å). In contrast, the D146 O $\delta$ 1 of wild-type PurR makes a close contact to the W147 O (2.9 Å), which is either stabilized by a shared proton or is an unfavorable interaction.

*Crystal structure of PurR W147A-hypoxanthine-purF operator and model of W147A corepressor binding domain.*

To understand more fully why the W147A PurR mutant binds hypoxanthine more tightly than wild-type PurR, and why W147A acts as a "super repressor" *in vivo*, we determined the structure of this mutant in the holo complex, with hypoxanthine and *purF* operator. Again, the corepressor binding pocket of the W147A-corepressor-*purF* operator complex revealed that none of the wild-type protein-purine corepressor interactions are missing and none are gained, nor does the purine take a radically different binding position. Superimposition of the C $\alpha$ s of this structure onto the

corresponding atoms of the wild type PurR-hypoxanthine-*purF* operator structure (1QPZ) confirms their structural identity and yields an rmsd of 0.23 Å (within the error of the coordinates). As observed in the wild type PurR-hypoxanthine-*purF* operator structure, the DNA is again kinked about the central CpG base pair step, with a DNA bending angle of 51°. All other DNA properties were also nearly identical to those found in the wild type protein-corepressor-DNA structure (data not shown).

The alanine mutation is considerably more drastic than the phenylalanine replacement, and a larger change in localized structure is observed (Figure 3.3C). The allosteric switch loop, residues 146-154, on which A147 resides is shifted toward Y126 and more mobile than the wild-type loop. This is evidenced by the higher average thermal factors in this region, 80 Å<sup>2</sup> for W147A compared to 66 Å<sup>2</sup> in wild-type. The rmsd between Cα carbons of wild-type and W147A is 0.89 Å for residues 146-154. The A147 Cα translocates 0.78 Å, but the most dramatic Cα displacement observed is for G148, 1.69 Å. The shift into the indole binding pocket fills only partially the gap created by the removal of the large tryptophan side chain. However, the Y126 side chain position is nearly identical in W147A and wild-type. The side chain of residue K151, on the other hand, moves toward Y126 to help fill the indole ring binding pocket. Also, because the bulky tryptophan has been removed, the K151 Cε-Nζ bond in the W147A structure is able to rotate to allow Nζ to make hydrogen bonds to the Oε1 and Oε2 atoms of residue E149, whereas in the wild-type structure only the K151 Nζ-E149 Oε1 hydrogen bond is observed.

Crystallization trials with the W147A-CBD were unsuccessful using standard conditions, and multiple *de novo* attempts have failed. In order to visualize the

corepressor binding pocket of apo W147A, a simple model was constructed (Figure 3D). The W147A substitution is clearly destabilizing, as a large cavity is created that renders the hydrophobic residues Y73 and F74 more solvent accessible, apparently an energetically costly change. Moreover, the wild-type W147 residue makes 29 contacts of less than 4.0 Å to atoms in surrounding residues, including hydrogen bonds to the Y73 hydroxyl and D160 carboxylate side chain and numerous van der Waals contacts. No contacts were made to the C $\beta$  of A147, supporting the hypothesis that the apo form of the W147A mutant is strongly disfavored.

*Crystal structure of PurR-W147R-hypoxanthine-purF operator and model of W147R corepressor binding domain.*

To elucidate the structural basis of the higher affinity of W147R for hypoxanthine and its "super repressor" phenotype, we determined the structure of the W147R-hypoxanthine-*purF* operator complex. Again, the corepressor binding pocket of the W147R-corepressor-*purF* operator complex revealed that none of the wild-type protein-purine interactions are missing, nor does the corepressor base take a radically different binding position. Superimposition of C $\alpha$ s of residues 3-340 of this structure onto the corresponding atoms of the wild type holo complex structure confirmed their structural identity and yielded an rmsd of 0.43 Å. The slightly higher rmsd for this structure is likely attributable to small global changes that resulted from collection of the data from a cryocooled crystal and comparing the resulting structure to the room temperature wild-type PurR structure. As in the wild-type PurR-hypoxanthine-*purF* operator structure, the DNA is again observed to be kinked about the central CpG base pair step, with a DNA

bending angle of 50°. All other DNA properties were also conserved relative to the wild type protein-corepressor-DNA structure (data not shown).

A tryptophan to arginine mutation is somewhat more conservative than an alanine replacement with respect to size and the ability to hydrogen bond, yet very similar *in vivo* repression activity is observed for both W147R and W147A (Table 3.1). The significant rearrangement of the indole ring binding pocket that is observed in the W147R mutant structure is larger than that observed in the other W147 mutant proteins (Figure 3.3A, C, E). Specifically, the entire allosteric switch loop, residues 146-154, is more flexible than the wild-type or W147F, as quantitated by the higher average thermal factors for C $\alpha$ s in this region, 81 Å<sup>2</sup>, compared to 66 Å<sup>2</sup> for wild-type and 65 Å<sup>2</sup> for W147F. Loop residues 146-149 move closer to Y126 relative to their positions in the wild-type structure, assuming positions similar to those observed in the structure of the other super repressor, W147A. The rmsd between the C $\alpha$  atoms of W147R and wild-type for residues 146-149 is 1.14 Å, while that between W147R and W147A in this region is 0.64 Å. Thus, W147R and W147A are structurally more similar to each other than either is to wild-type PurR. However, some differences are found between W147R and both W147A and wild-type PurR. Residues 150-154 shift away from the R147 and Y126 side chains, unlike the indole pocket filling movement seen in the W147A structure. However, the Y126 side chain position is conserved. In fact, the guanidinium group of the R147 side chain makes extensive contacts (18 less than 4.0 Å) with the tyrosine ring, including a cation- $\pi$  interaction. The two closest approaches of the cation- $\pi$  interaction are between R147 N $\epsilon$  and Y126 C $\gamma$  (3.0 Å) and C $\delta$ 2 (3.0 Å). The K151 side chain is thermally diffuse in this

structure but has moved farther from the R147 relative to its wild-type position. There is an additional water molecule found in the place of the wild-type tryptophan C $\zeta$ 2.

Crystallization trials with the W147R-CBD were also unsuccessful using our standard wild-type condition. As with the W147A-CBD, no alternative crystallization reagent has proven fruitful. Hence, we constructed a model of the apo corepressor binding pocket of the W147R mutant. When the R147 is modelled to lay flat in the plane occupied by W147 (in magenta, Figure 3.3F), a hydrogen bond analogous to that found in wild-type can be made to the hydroxyl group of Y73. However, to allow this hydrogen bond to form, the arginine must push its guanidinium group N $\zeta$ 1 and N $\zeta$ 2 atoms deep into a hydrophobic cleft of the corepressor binding pocket, which has no suitable hydrogen bonding partners nearby. The wild-type W147 makes hydrophobic contacts in this cleft with I77, M122, M145, V158, and L297. An alternative and seemingly better model was constructed by allowing the R147 side chain to rotate and thereby allow its N $\epsilon$  to hydrogen bond with D160 (light blue, Figure 3.3F). However, this leaves the corepressor binding pocket open, and a cavity is formed (similar to that of the modelled W147A-CBD).

## Discussion

The crystal structures of the purine repressor in its corepressor bound and free conformations have shown that differences in intersubunit interactions, brought about by movement of a "switch" loop, are responsible for transducing the intramolecular signal from the corepressor binding domain (CBD) to the DNA binding domain (DBD) that corepressor has bound and thus allowing the DBD to bind DNA (Schumacher et al., 1995; Schumacher et al., 1994a). The C-terminal subdomain of the CBD moves very little, while the N-terminal subdomain of the CBD rotates  $\sim 20^\circ$ , bringing the hinge helices into contact with one another and allowing the hinge region of the DBD to become structured. This is similar, but not identical, to that observed for the lactose repressor, LacI, which has a remarkably similar fold, but in which the N-terminal subdomain rotates only  $6^\circ$  and is translated  $\sim 4 \text{ \AA}$  (Bell and Lewis, 2000; Lewis et al., 1996). However, LacI represses transcription of catabolic genes and so represses only in the absence of a ligand, allolactose. In contrast, PurR binds either guanine or hypoxanthine as a corepressor to then repress the transcription of genes of the *pur* regulon (involved in the biosynthesis of purines). In spite of this difference, the mechanism of rotating the N-terminal subdomain of the CBD to alter the relative position of the hinge helices is conserved between the two repressors.

The CBD and DNA-bound structures of PurR reveal the end states of its allosteric transitions. To delineate the roles of specific residues on this allosteric mechanism, we have employed biochemical and structural approaches on one key residue, W147. In initial studies on corepressor binding to PurR, residue W147 was hypothesized to interact with the guanine or hypoxanthine corepressor directly because mutation to W147F

resulted in weaker *in vivo* repression and lower affinity for hypoxanthine and corepressor (Choi et al., 1994; Schumacher et al., 1993). Subsequently, the crystal structures of PurR in its holo and apo forms revealed that W147 is involved, but surprisingly this residue makes no contacts to the corepressor as postulated (Schumacher et al., 1995; Schumacher et al., 1994a). Rather, W147 has a dual role whereby in one case the indole ring acts as a structural but not functional surrogate of corepressor that helps stabilize the "open", nonrepressing conformation of PurR by occupying the corepressor binding pocket. In contrast, when a corepressor is present, W147 translocates to an "indole binding pocket", which helps to maintain the "closed", repressing conformation. To explore the functional importance of the indole side chain at position 147, three substitutions, W147F, W147A, and W147R, were made and characterized *in vivo* and *in vitro*. Crystallographic studies were undertaken to reveal the stereochemical and structural parameters that contribute to the biological and biochemical functions of each mutant.

The W147F mutant is a weaker physiological repressor (Table 3.1). This correlates well with *in vitro* corepressor binding data, in which W147F binds corepressor 15-fold less well than wild-type PurR. However, in the presence of excess hypoxanthine, the DNA binding function of W147F is unimpaired (Table 3.1). The structure of the apo W147F protein shows that the phenyl ring engages in protein-style stacking interactions with the side chains of Y73 and F74. The loss of the hydrogen bond between the Nε of W147 and Y73 hydroxyl is the likely cause of the higher mobility of the latter side chain. This gain in side chain entropy appears to offset at least somewhat the enthalpic loss as far as the relative stability of the W147F mutant compared to wild-type PurR. The loss of corepressor affinity appears to lie in the corepressor-bound form. When hypoxanthine

binds and the phenylalanine side chain has translocated from the corepressor binding pocket, the smaller phenyl ring is unable to stack as tightly against Y126. Moreover, a small cavity is created, which is energetically unfavorable (Matthews, 1996; Rashin et al., 1997; Vlassi et al., 1999).

The W147A phenotype is that of a super repressor (Table 3.1). *In vitro* data show that this phenotype is largely due to the at least thirty-fold increase in affinity for corepressor. The W147A mutant is structurally destabilized in the corepressor-free form because the alanine cannot form a hydrogen bond with Y73 or make stacking and van der Waals interactions with Y73, F74, and other residues with which the tryptophan indole ring makes contacts. Also, the indole binding pocket is normally a solvent-accessible hydrophobic patch, with interior hydrophobic residues such as M145 and V158 shielded from bulk solvent by W147. The uncovering of these residues that results from the W147A substitution, taken together with the energetic cost of creating a cavity, provide a plausible explanation for the sharp increase in affinity for corepressor. Loss of the W147 side chain may also make the corepressor binding pocket more accessible to corepressor. W147A may be structurally more similar to the "closed" conformation even in the absence of corepressor, or shifted toward that conformation more often in the equilibrium of open and closed forms. This is supported by the higher affinity of W147A for *purF* operator than wild-type PurR (0.7 +/- 0.2 nM and 2.5 +/- 0.3 nM, respectively). There is little effect on the structure of the holo complex. The allosteric switch loop, residues 146-154, closes in on the cavity created by removal of the tryptophan, and we see the formation of a new hydrogen bond between the N $\zeta$  of K151 and the E149 O $\epsilon$ 2, a bond that is sterically disfavored in wild-type PurR. It is possible that the W147A mutant is



stabilized in the "closed" form to such a degree that it is unable to enter the "open" form, thus explaining our inability to obtain corepressor-free W147A-CBD crystals. Therefore, we conclude that the conformational destabilization of the apo form makes this mutant a super repressor.

As demonstrated by our *in vivo* repression experiments, W147R is also a super repressor. Those studies are extended by our *in vitro* binding data, which demonstrate that W147R binds hypoxanthine with two-fold higher affinity than wild-type, whilst cognate DNA binding affinity is unaffected. In the W147R-hypoxanthine-*purF* operator structure, we see that the R147 side chain moves closer to the Y126 side chain, and the entire side chain beyond C $\beta$  makes energetically favorable contacts with Y126, including van der Waals interactions and a cation- $\pi$  interaction with the aromatic ring of Y126 (Wouters, 1998). The apo form of W147R is likely to be destabilized by the inability of the arginine side chain to make stacking interactions with the planar residues in the corepressor binding pocket. Our modelling studies suggest that the long, electropositive arginine side chain might not be accommodated in the pocket, and its less hydrophobic nature surely disfavors its positioning there as a corepressor substitute. We constructed two models for R147 interactions in the corepressor binding pocket in which we optimized hydrogen bonding and avoided steric clash (Figure 3.3). Each model revealed obvious problems with fitting the R147. Either a large cavity was formed, as in the W147A-CBD model, or a positive charge was buried in the midst of hydrophobic residues, both of which are energetically costly. The difficulty of accommodation of R147 in the "open" conformation without restructuring of the apo corepressor binding

pocket, along with the propitious interactions between R147 and Y126 in the "indole" binding pocket, account for the super repressor activity of the PurR-W147R mutant.

In conclusion, we have determined binding affinities and crystal structures of three allosteric signalling mutants, PurR W147F, W147A, and W147R. Mutations at position 147 have a significant effect on *in vivo* repression activity and *in vitro* corepressor binding, although this residue does not contact corepressor directly. This is a key residue in the allosteric response mechanism of PurR and it plays a role not unlike the intrasteric autoregulatory sequences of many allosteric proteins (Kobe and Kemp, 1999). The crystal structures of each of these mutants give insight into the complexity of the interactions necessary for intramolecular signal transduction. Interestingly, LacI-inducer and LacI-DNA complexes do not reveal a similar "switch" in transiting between ligand and DNA-bound conformations (Bell and Lewis, 2000; Lewis et al., 1996). It is possible that the movement of W147, and subsequently the entire "switch" loop, is linked to the larger rotation of the N-terminal subdomain of the CBD (20°), relative to that observed for LacI (6°). Perhaps other LacI/GalR family members, such as the *E. coli* ribose repressor, RbsR, that share a tryptophan at this position will undergo similarly large rotations, whilst members that do not have a tryptophan at this position will undergo some alternative movement. Notably, GalR has an arginine at this position, corresponding to one of our super repressor mutants, however no structural or mutant binding data is yet available for GalR. The inducer bound structure of the trehalose repressor, TreR, has been solved (Hars et al., 1998). This structure reveals that an arginine at position 147, which correlates in a structure-based sequence alignment with W147 of PurR, directly contacts the hydroxyl groups at the exocyclic 2 and 3 positions of

both trehalose and trehalose-6-phosphate. It is possible that the differing mechanisms of repression of PurR and other LacI/GalR family members, i.e., binding a small molecule corepressor or inducer, accounts for the altered use of residues on this loop. Structural information from more family members in both their repressing and nonrepressing conformations will be required to determine whether the "switch" loop mechanism is unique to PurR.

### **Acknowledgments**

The *purR-W147F*, *A*, and *R* mutant plasmids were provided by the Zalkin laboratory. The *in vivo* repression assays were performed by Kang-Yell Choi and Fu Lu of the Zalkin laboratory. I carried out all *in vitro* equilibrium binding experiments and the determination of all of the structures presented in this chapter.

Table 3.1. *In vivo* repression and *in vitro* equilibrium binding of wild-type PurR and W147 mutants as determined by fluorescence polarization

	<i>In vivo</i> repression, fold	$K_d$ ( <i>purF</i> ), nM	$K_d$ (hyp), $\mu$ M
Wild-type	13.9 <sup>A</sup>	2.5 +/- 0.3	1.6 +/- 0.3
W147F	4.9 <sup>A</sup>	2.2 +/- 0.3	24.0 +/- 4.7
W147A	31.3 <sup>B</sup>	0.7 +/- 0.2	< 0.05
W147R	33.6 <sup>C</sup>	2.0 +/- 0.4	0.85 +/- 0.25

<sup>A</sup> These results were reported previously in (Choi et al., 1994).

<sup>B</sup> This result was reported in (Lu et al., 1998a).

<sup>C</sup> This result was reported in (Lu, 1999).

Table 3.2. Data collection and refinement statistics for W147F corepressor binding domain and mutant PurR-hypoxanthine-*purF* operator complexes

	W147F-CBD	W147F-hyp- <i>purF</i>	W147A-hyp- <i>purF</i>	W147R-hyp- <i>purF</i>
<i>Data collection:</i>	SSRL <sup>††</sup>	ADSC <sup>†</sup>	ADSC <sup>†</sup>	SSRL <sup>††</sup>
Resolution (Å)	2.40	2.90	2.50	2.55
Unique reflections (#)	20934	13773	23194	22093
Completeness (%)	93	91	97.5	96
<i>I</i> /sig( <i>I</i> ) for data from				
11.5 to 2.0 Å	6.64			
10 to 2.9 Å		4.29		
11.5 to 2.5 Å			7.55	
∞ to 2.55 Å				4.89
R <sub>sym</sub> (%)	7.9	7.54	6.4	11.2
Temperature (°C)	25	25	25	-180
Cell constants:				
a (Å)	38.02	176.05	175.92	175.38
b (Å)	126.13	95.04	95.25	92.61
c (Å)	61.84	81.51	81.35	80.64
α (°)	90	90	90	90
β (°)	99.71	90	90	90
γ (°)	90	90	90	90
<i>Refinement:</i>				
Resolution (Å)	2.4	2.9	2.5	2.55
R factor (%)	17.8	15.7	19.1	21.9
Total atoms (#)	4489	3040	3085	3121
Water molecules (#)	151	32	89	107
rms deviations:				
bond length (Å)	0.009	0.016	0.017	0.016
bond angle (°)	1.018	1.757	1.745	1.655
B correlation (Å <sup>2</sup> )	2.662	4.10	4.619	4.604

<sup>†</sup> ADSC = Area Detector Systems Corporation

<sup>††</sup> SSRL = Stanford Synchrotron Radiation Laboratory

Figure 3.1. Illustrations of the corepressor binding domain from A. the holo PurR-hypoxanthine-*purF* operator structure, and B. the apo corepressor binding domain structure. Trp147 atoms are shown as space-filled models, in pink. Close-up views of the corepressor binding pocket in the: C. PurR-hypoxanthine-*purF* operator complex structure and D. corepressor free PurR CBD structure. Holo complex structures are shown in blue, and apo CBD structures are shown in aquamarine.

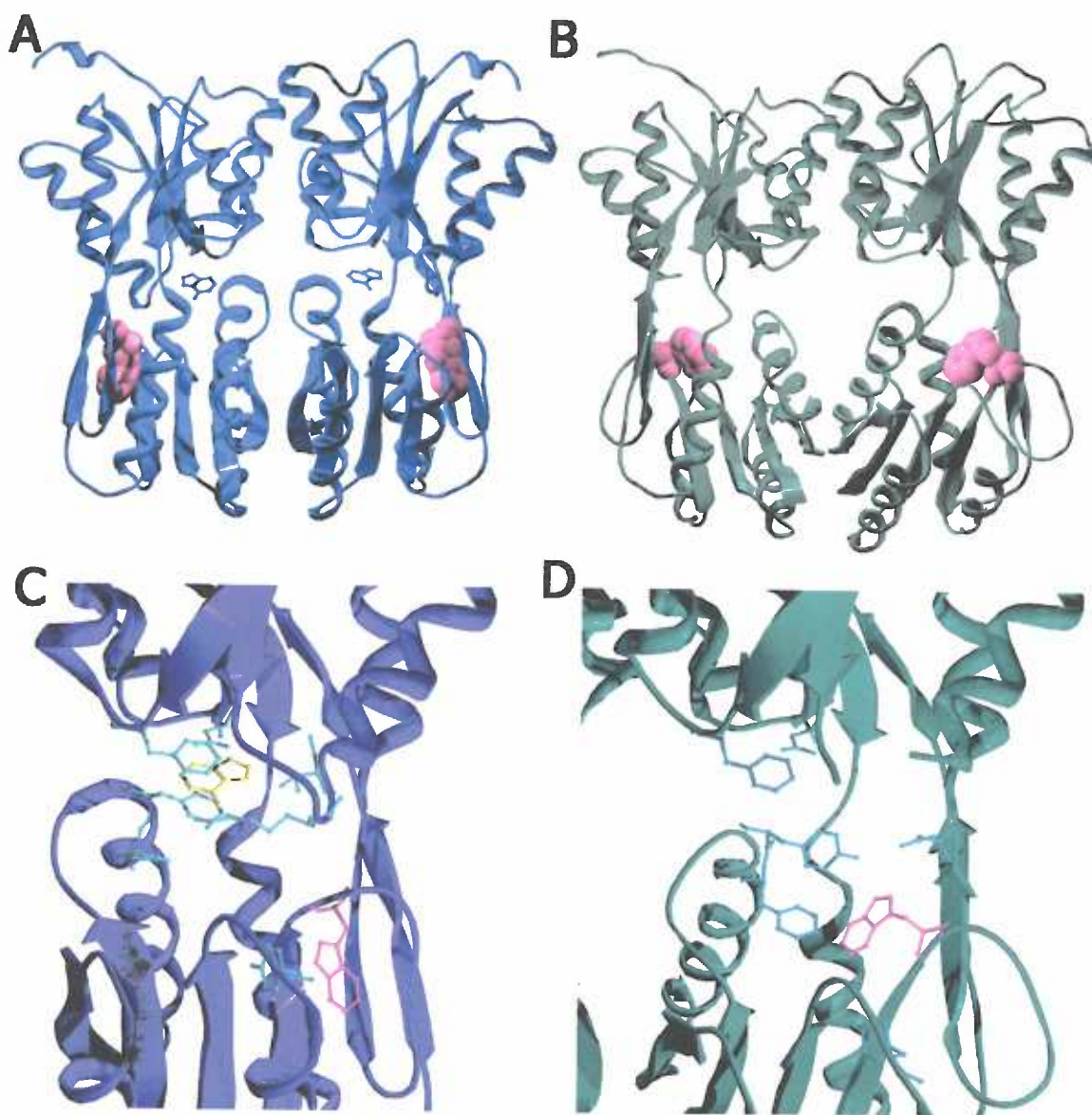


Figure 3.2. Representative binding isotherms for wild-type and W147 mutant proteins.

A. Binding to DNA in the presence of saturating amounts of corepressor. Symbols represent experimental values, and lines are curve fits to those data: wild-type PurR (●, —), W147F (+, ---), W147A (◇, --), W147R (□, - -). B. PurR binding of corepressor as measured indirectly through DNA binding. Symbols and lines are the same as used in A.



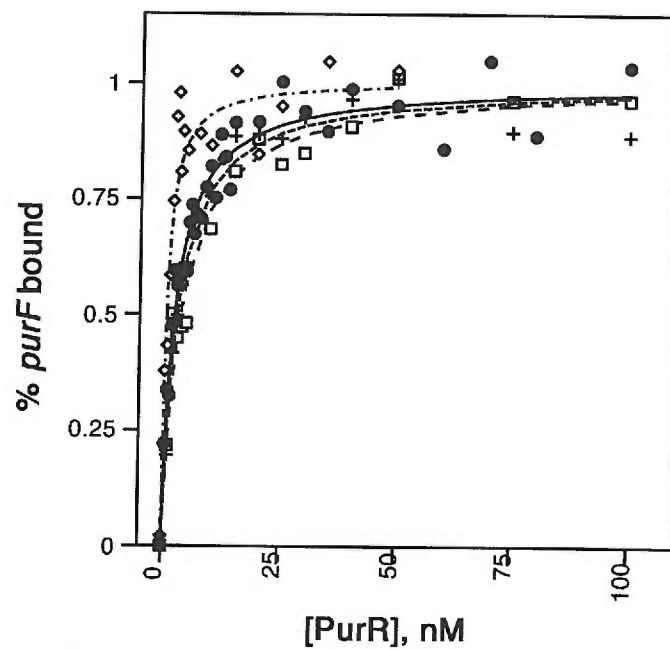
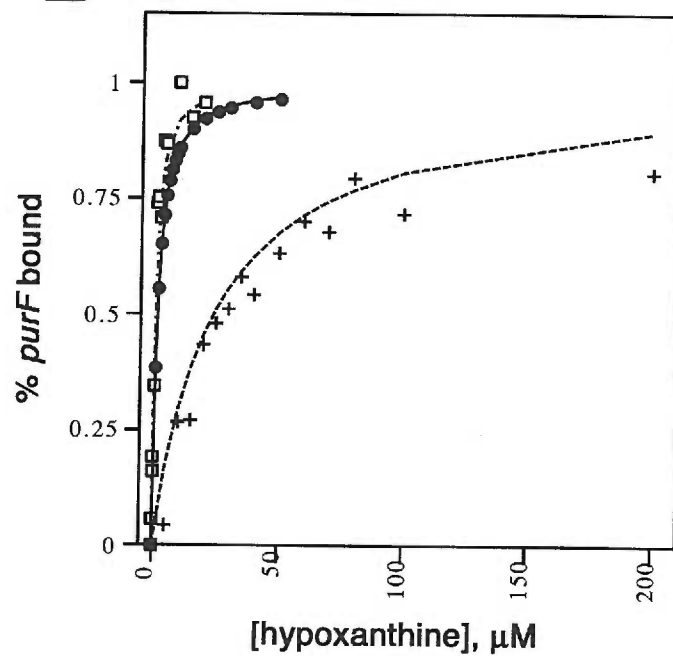
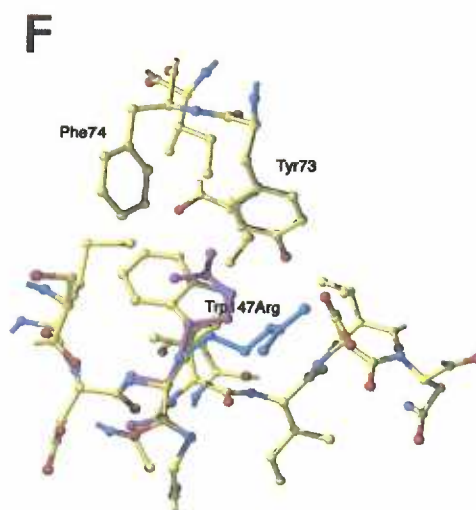
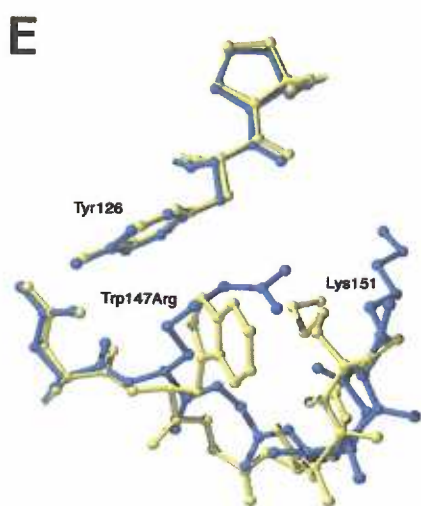
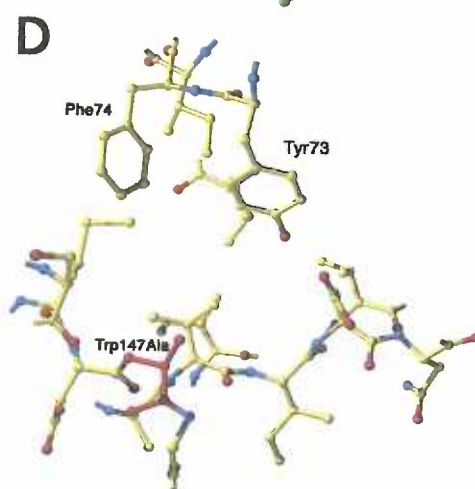
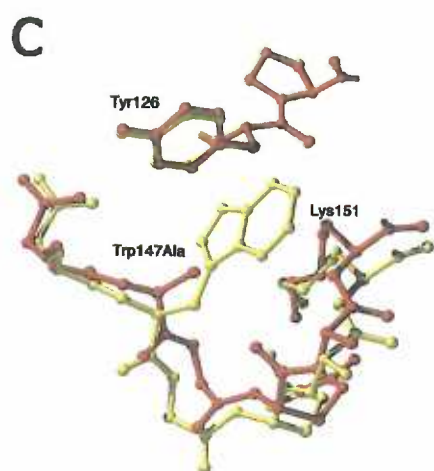
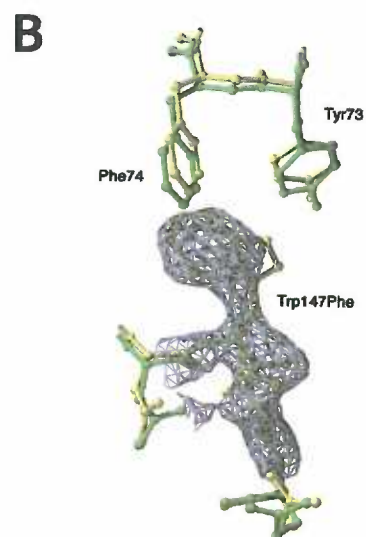
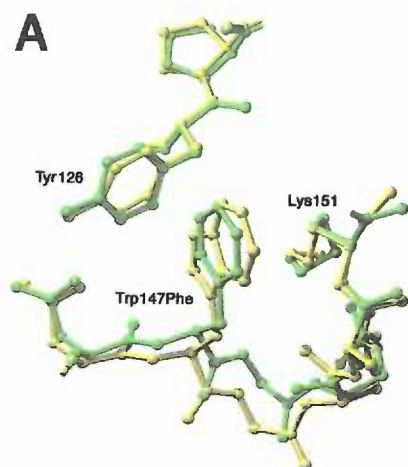
**A****B**

Figure 3.3. Overlays of the corepressor binding pockets of each of the mutant structures solved onto the wild-type structure. A. W147F-hypoxanthine-*purF* operator, in green, overlaid onto the wild-type structure, in yellow. B. W147F-CBD is shown in green, wild-type CBD in yellow.  $2F_o - F_c$  electron density, contoured at  $1.5 \sigma$ , is shown to illustrate the *cis* peptide bond between residues 147 and 148. C. W147A-hypoxanthine-*purF* operator, in red, and wild-type in yellow. D. Model of the W147A-CBD corepressor binding pocket, with the alanine shown in red. E. W147R-hypoxanthine-*purF* operator, in blue, and wild-type in yellow. F. Two possible positions for the W147R-CBD R147, one shown in light blue and the other in magenta.



**Chapter 4. Structure-based expansion of corepressor specificity of the *Escherichia coli* purine repressor by substitution of residue 222**

Joy L. Huffman<sup>1</sup>, Maria A. Schumacher<sup>1</sup>, Dennis N. Arvidson<sup>1</sup>, Fu Lu<sup>2</sup>, Howard Zalkin<sup>2</sup>,  
and Richard G. Brennan<sup>1</sup>

<sup>1</sup>*Dept. of Biochemistry and Molecular Biology, Oregon Health Sciences University*

<sup>2</sup>*Department of Biology, Purdue University*

repressor by making a similar water-mediated contact between the side chain of Lys222 and the exocyclic oxygen at the 2 position of xanthine. These results establish at the atomic level the basis for the critical role of Glu222 in the recognition of the exocyclic 2 position of guanine and for the successful expansion of corepressor specificity.

## Introduction

The *Escherichia coli* purine repressor (PurR) binds to a conserved 16 base pair operator sequence and regulates transcription of multiple *pur* regulon genes, including nine operons for *de novo* purine nucleotide biosynthesis, the *purR* gene and other genes with functions related to nucleotide metabolism (Zalkin and Nygaard, 1996). PurR is a 341 amino acid residue polypeptide that functions as a dimer. Each subunit is composed of an N-terminal DNA binding domain (residues 1-60) and a larger C-terminal corepressor binding and dimerization domain (residues 61-341), termed the CBD. Binding of a corepressor, hypoxanthine or guanine, increases the operator DNA binding affinity. Consequently, PurR is the primary regulatory protein that coordinates expression of *pur* regulon genes to the availability of purines in *E. coli*.

Comparison of the corepressor binding pockets of PurR bound to hypoxanthine and guanine implicated residue Glu222 as a key specificity determinant of PurR against 2-oxopurines and explained the preference of PurR for guanine over hypoxanthine (Figure 4.1) (Schumacher et al., 1997). Specifically, the carboxylate group of this residue and the carbonyl oxygen of Phe221 were found to accept hydrogen bonds from a water molecule, thus leaving the water molecule with only hydrogen bond acceptors and resulting in a water-specific interaction with the hydrogen bond donating N2 nitrogen atom of guanine. Such hydrogen bond complementarity should preclude PurR binding to any purine that has hydrogen bond acceptor groups at this position, e.g., xanthine (Figure 4.2). In addition, the E222 carboxylate group contributes to the electronegative environment about the 2 position of corepressor providing additional corepressor

discrimination for the more electropositive N2 of guanine and against the more electronegative O2 of xanthine.

In this report we describe the structure-based design and analysis of three mutant PurR proteins in which residue Glu222 is substituted by alanine, glutamine, or lysine. To study the function of mutations *in vivo*, we used conditionally replicative, integrative plasmids to incorporate *purR* into the chromosome in single copy. Next we purified the mutant proteins to assess their corepressor and DNA binding activities. Finally, we describe nine crystal structures of mutant PurR-corepressor-*purF* operator complexes in which Glu222 has been replaced by alanine (E222A), glutamine (E222Q), or lysine (E222K). Replacement of Glu222 by alanine or by glutamine reduced the preference of PurR for guanine over hypoxanthine. The structures of E222A and E222Q bound to hypoxanthine or guanine reveal the structural and chemical features that allow these purines, which differ at their exocyclic 2 position, to bind the repressor with nearly equal, albeit lowered, affinity. The structures of E222K bound to hypoxanthine, guanine or xanthine clarify the stereochemical details of the binding interactions with the natural corepressors and show the water-mediated interactions which this protein uses to bind xanthine with only 1.2 or 2 fold lower affinity than hypoxanthine and guanine, respectively. Combined, the *in vivo*, *in vitro* and structural data demonstrate that Glu222 is the key recognition determinant of PurR that gives rise to preference for 2-aminopurine corepressors and prevents tight binding of 2-oxopurine corepressors.

## Materials and Methods

### *Strains and plasmids.*

*E. coli* strains and plasmids have been described (Lu et al., 1998b). Plasmid pCAH64 is a chloramphenicol-resistant, *oriR*<sub>6Kγ</sub> *attP* plasmid that was derived from pSK49Δ*uidA*2 as described (Haldimann et al., 1996). In pCAH64 *purR* is transcribed from a weak synthetic promoter denoted *P*<sub>synI</sub> that originated from pHH7013 (from J. Hu). Similar plasmid constructs in which *purR* was expressed from *P*<sub>tac</sub> instead of *P*<sub>synI</sub> were toxic, presumably due to high expression of *purR* (Rolfes and Zalkin, 1988a). The mutant *purR* genes were subcloned into a precursor plasmid pCAH63 as *NdeI*/*Bam*HI fragments, thus yielding pCAH64.

All *pir*-dependent plasmids were propagated in the *pir*<sup>+</sup> host BW23473 or the *pir*-116 host BW23474 as described (Lu et al., 1998b). Plasmids were integrated, in single copy, into the chromosomal *attB* site in strain FL100 using pINT-ts as described (Lu et al., 1998b).

### *Assay for repression by PurR.*

FL100 *purR* integrants were cultured overnight at 37° C in minimal medium (Rolfes and Zalkin, 1990b) containing 15 μg/ml chloramphenicol. Then 100 μl of the fresh overnight culture was used to inoculate 5 ml of the same medium in a 1.5 x 12 cm tube. The cells were grown with shaking at 37° C until growth reached mid-log phase. The β-galactosidase activity was assayed in permeabilized cells according to the method of Miller (Miller, 1972).



### *Site-directed mutagenesis.*

Site-directed mutagenesis by the procedure of Kunkel *et al.* (Kunkel, 1985) was carried out as described previously (Choi *et al.*, 1994). Single-stranded *purR* template DNA was prepared from plasmid pPR1005-2 (Rolfes and Zalkin, 1990b).

### *Overexpression and purification of PurR.*

Wild-type and mutant *purR* genes in pPR1005-2 were transferred via flanking *NdeI* and *HindIII* sites to pET24(a) for overexpression. Overexpressed PurR was purified as described (Choi and Zalkin, 1992). For purified PurR, the protein concentration was determined spectrophotometrically at 280 nm using  $A_{280}$  (1%) of 12.9 (Rolfes and Zalkin, 1990b).

### *Corepressor binding.*

Equilibrium dialysis was used to determine binding of purine corepressors when the  $K_d$  for guanine, hypoxanthine or xanthine was less than 20  $\mu$ M. The binding measurements were carried out as described previously (Choi and Zalkin, 1992). Binding data were fit to the Hill equation using Ultrafit software (Biosoft, Cambridge, UK) (Eq. 1). Y is the fraction of protein bound;  $K_d$ , the dissociation constant; n, the Hill coefficient; L, free corepressor. Fits were made allowing both  $K_d$  and n to vary.

$$Y = \frac{[L]^n}{K_d^n + [L]^n} \quad (1)$$

In cases where the  $K_d$  values were above 20  $\mu\text{M}$ , gel retardation was used to estimate binding of corepressors to PurR (Rolfes and Zalkin, 1990b). In this method, corepressor binding to PurR is coupled with binding of the holorepressor to operator DNA. The method relies on the observation that the affinity of the PurR-corepressor complex for DNA is more than 200 fold greater than the affinity of PurR for corepressor (Choi et al., 1994; Choi and Zalkin, 1992; Rolfes and Zalkin, 1990b). This relationship assures that once corepressor binds to PurR the complex will bind to DNA. A 30-mer DNA probe with the *purF* operator was labeled with  $^{32}\text{P}$  as described (Choi et al., 1994). Binding was performed in corepressor-dependent Buffer II (100 mM potassium HEPES, pH 7.5, 250 mM potassium glutamate, 150 mM NaCl, 10 mM magnesium acetate, 1 mM EDTA, 2% dimethyl sulfoxide). The 20  $\mu\text{l}$  incubation contained 10 fmol DNA probe, 15 nM PurR, varied concentrations of corepressor and Buffer II. The concentration of PurR refers to the 38 kDa subunit. The protein-DNA complex and free DNA were separated by electrophoresis and the gel was dried and counted for radioactivity using a Packard Instant imager. From these data the fractional saturation of PurR was estimated at varied concentrations of corepressor. Previous control experiments confirm that the two methods give similar  $K_d$  values for the PurR-hypoxanthine interaction (Lu et al., 1998b).

#### *Measurement of apo PurR-DNA binding.*

The specific binding of wild-type and mutant apo PurR to operator DNA was determined by gel retardation using corepressor-independent Buffer I (10 mM HEPES-KOH, pH 8.0, 100 mM KCl, 1 mM EDTA) (Rolfes and Zalkin, 1990b). In this buffer

PurR assumes the conformation required for high affinity binding to operator DNA without a requirement for corepressor. This assay was used to evaluate the DNA binding function of corepressor binding mutants.

*Crystallization, data collection and structure determination.*

Crystals of the mutant PurR-purine-*purF* operator complexes were grown by the hanging drop-vapor diffusion method, which was used previously to crystallize the PurR-hypoxanthine-*purF* operator complex (Schumacher et al., 1994b). Briefly, protein at a concentration of 0.25 to 0.50 mM was saturated with corepressor and mixed with a 0.80 mM solution of the 16 base pair *purF* operator in a 1:1.5 molar ratio (Glasfeld et al., 1996; Schumacher et al., 1994b). The protein-corepressor-DNA solution was mixed 1:1 with the reservoir solution, which was 25% polyethylene glycol (PEG) 4000, 0.40 M ammonium sulfate, 50 mM cobalt hexamine chloride and 0.10 M ammonium phosphate, pH 7.5.

Crystals grow initially as thin two-dimensional plates and, within a period of two weeks to several months, melt and are replaced by large rhombohedral crystals (typically 0.6 mm x 0.3 mm x 0.2 mm). The mutant PurR-purine-*purF* operator crystals were isomorphous to the wild-type PurR-hypoxanthine-*purF* operator crystals (Table 2). X-ray intensity data were collected at room temperature with an R-AxisIV (MSC) image plate detector using a Rigaku RU300 rotating anode operating at 50 kV and 100 mA. Selected crystallographic data for each mutant-purine-DNA complex are listed in Table 4.2.

### *Structure determination and refinement.*

All nine structures were solved by difference Fourier techniques using the wild-type PurR-hypoxanthine-*purF* operator complex (1QPZ), after removing the corepressor, the glutamate side chain of residue 222, and all solvent, as the starting model for each (Schumacher et al., 1994a). The asymmetric unit contained one PurR monomer, one purine molecule, and a *purF* operator half-site. Refinement of each mutant complex structure began with several cycles of rigid body refinement and was followed by XYZ (positional) and B factor (thermal parameter) refinement as implemented in TNT (Tronrud et al., 1987). Purines and water molecules were built into the structures in the final stages of refinement using either FRODO (Jones, 1985), O (Jones, 1991), or Xfit (McRee, 1992). To ensure the unbiased placement of the purines and any side chains that underwent significant movement, a series of  $F_o - F_c$  omit maps were calculated for each structure. In these omit map calculations, the purine, attendant water molecules, and any rebuilt corepressor binding pocket residues were removed. The model was then subjected to multiple cycles of positional and thermal refinement before inspection of the omit density map. Refinement statistics and selected stereochemical parameters (Laskowski et al., 1993) are listed in Table 4.2.

## Results

### *In vivo function of PurR mutants.*

Based on the critical interaction of Glu222 with the purine corepressor exocyclic 2 position, as revealed by the X-ray structures of PurR bound to operator DNA (Schumacher et al., 1994a; Schumacher et al., 1997), we mutagenized this residue to alanine, glutamine, or lysine. Mutant genes were constructed and integrated into the phage  $\lambda$  attachment site of the *E. coli* chromosome in single copy in order to evaluate repressor function. The *purR* gene in this construction is controlled by a weak synthetic promoter chosen to give expression comparable to the wild-type gene at its normal site (Lu et al., 1998b). Western blots of wild-type cells and the integrants verified that all strains contained similar PurR levels (not shown). Repression of the *purF-lacZ* reporter by wild-type PurR, E222A, E222Q, and E222K was 18.0, 8.5, 7.6 and 4.3 fold, respectively (He et al., 1990; Rolfes and Zalkin, 1988b). We note that the repression assay depends upon corepressor-dependent binding of PurR to the *purF* operator but does not address which purines function as corepressor.

### *Purification of wild-type and mutant PurR proteins.*

The wild-type and E222A, E222Q, and E222K mutant proteins were overexpressed as described in the Materials and Methods, and the repressor protein was purified in order to assess corepressor binding directly. For wild-type and the three mutants, PurR accounted for approximately 40% of the soluble protein based on sodium dodecylsulfate polyacrylamide gel electrophoresis (not shown). The elution profiles of

PurR from DEAE Sepharose and heparin agarose columns were identical for wild-type and mutants. In addition, similar binding constants to operator DNA were obtained for the wild-type and mutant repressors when determined in corepressor-independent binding Buffer I (Rolfes and Zalkin, 1990b). The apparent  $K_d$  for wild-type PurR was 11 nM and the values for the mutants were between 7 and 16 nM (data not shown).

#### *Binding of purines to PurR.*

Two methods were used to determine binding affinities of purines to wild-type and mutant repressors. Equilibrium dialysis was used to determine binding of hypoxanthine and guanine in cases where the  $K_d$  was lower than 20  $\mu$ M. Due to the limited solubility of PurR in dialysis buffer at 4° C, binding constants were estimated by a gel retardation assay in cases of low affinity (Rolfes and Zalkin, 1990b). Binding constants are summarized in Table 4.1. For wild-type PurR,  $K_d$  values of 1.6  $\mu$ M and 11  $\mu$ M were obtained for guanine and hypoxanthine, respectively, in close agreement with previous data (Choi and Zalkin, 1992). The apparent  $K_d$  value of wild-type PurR for xanthine was 233  $\mu$ M. Thus, the exocyclic atom at position 2 of the purine (a hydrogen bond donor for guanine (N2) and a hydrogen bond acceptor for xanthine (O2), as seen in Figure 4.2) is critical for the increased binding affinity of guanine over xanthine by a factor of more than 140.

#### *Binding of purines to Glu222 mutants.*

To evaluate the proposed key role of Glu222 for corepressor specificity, we determined binding affinities of purines for the E222A, E222Q, and E222K mutants.  $K_d$  values summarized in Table 4.1 show that for the E222A repressor, the binding affinity for guanine and hypoxanthine was decreased by 8 and 2.5 fold, respectively, whereas the binding affinity for xanthine increased by 1.3 fold as compared to the wild-type. These changes in binding constants reduce the preference for binding guanine over hypoxanthine from 6 fold to 2.1 fold. Similar effects on specificity for corepressor were observed for the E222Q repressor. For the E222K mutant, the affinity for guanine and hypoxanthine decreased by 13.7 and 3.4 fold, respectively, while the affinity for xanthine increased by 5.2 fold as compared to wild-type PurR. Thus, the preference for hypoxanthine over xanthine was reduced from 21 fold to 1.2 fold and for guanine over xanthine from 146 fold to 2 fold. This brings the  $K_d$  for xanthine, 45  $\mu$ M, within a physiologically relevant range, as it is near the  $K_m$  of the *E. coli* xanthine phosphoribosyltransferase, ~30  $\mu$ M (Vos et al., 1997).

*Crystal structures of E222A bound to the purF operator and corepressors hypoxanthine, guanine, or xanthine.*

To determine the structural basis of E222A binding to the physiological corepressors, hypoxanthine and guanine, and its increased affinity for xanthine, we determined crystal structures of the E222A-hypoxanthine-*purF* operator complex, the E222A-guanine-*purF* operator complex, and the E222A-xanthine-*purF* operator complex. Superimposition of the C $\alpha$  carbons of these structures and the wild-type PurR-

hypoxanthine-*purF* operator structure confirms that they are identical with root mean-squared deviations (rmsds) of 0.25 Å for the E222A-hypoxanthine-*purF* operator complex, 0.22 Å for the E222A-guanine-*purF* operator complex, and 0.20 Å for the E222A-xanthine-*purF* operator complex, all of which are within error of the coordinates. As in the wild-type PurR-DNA structure, the DNA is kinked about the central CpG base pair step. The DNA bending angle was calculated with CURVES (Lavery and Sklenar, 1988) and excluded the first and last base pairs of the 16 base pair operator because of their poor electron density. This bending angle is 55° for the structures in which E222A is bound to hypoxanthine or xanthine and 47° for the E222A-guanine-*purF* operator structure. These values indicate that the DNA conformation is essentially identical to that of wild-type PurR-hypoxanthine-*purF* operator (Schumacher et al., 1994a). All other DNA properties, e.g., roll and twist angles and rise per nucleotide, were also nearly identical to those found in the wild-type protein-corepressor-DNA structure (Schumacher et al., 1994a).

To explain the moderately expanded specificity and the equilibrium dissociation binding constants of E222A for these three purines, we predicted that all other protein-corepressor contacts would be maintained and water molecules would be recruited to the binding pocket to replace the missing Glu222 interactions with protein-water-corepressor contacts. Xanthine was expected to bind better after removal of the negative electrostatic contribution of the Glu222 side chain. Examination of the corepressor binding pockets of the structures of the E222A-corepressor-*purF* operator complexes reveals, as expected, that none of the remaining wild-type protein-purine interactions are missing nor do the



purines take radically different binding positions (Figure 4.3A). These include aromatic interactions between the purines and Tyr73, Phe74 and Phe221, van der Waals interaction with C8, a water-mediated hydrogen bond from Arg190 to the exocyclic O6, a hydrogen bond to N7 from Thr192 side chain, and a hydrogen bond between N9 and the carboxylate group of Asp275.

Water-mediated contacts and the removal of protein-purine electrostatic interference explain the basis for the slightly increased affinity of E222A for xanthine over wild-type PurR. Because the exocyclic 2 positions of hypoxanthine and guanine are both electropositive (Figure 4.2) the wild-type Glu222 side chain electrostatically complements those bases and disfavors binding of xanthine. The mutation of Glu222 to alanine effectively removes this electronegative contribution to the binding pocket. In the E222A structures, the water networks around the purine bases are conserved, however the roles of each water, namely as hydrogen bond donors or acceptors, have changed to accommodate xanthine. Specifically, three waters found in the hypoxanthine-bound wild-type PurR structure (W1-3, Figure 4.3A) are visualized in the E222A-guanine-DNA structure. Only one of these water molecules, W1 in Figure 4.3A, is observed in the E222A-hypoxanthine-DNA structure, although this difference may be due to the low resolution of the data (2.7 Å). The E222A-xanthine-DNA structure, however, does contain the three waters close to their wild-type PurR-hypoxanthine positions. In the wild-type PurR-hypoxanthine-DNA structure, water molecule W3 donates a hydrogen to bond to the purine N3 and another to the backbone carbonyl oxygen of Phe221. In the E222A-xanthine-DNA structure, this water is conserved but shifted slightly to allow a

weaker hydrogen-accepting interaction with the N3 of xanthine and a closer, hydrogen bond-donating interaction with the xanthine O2. Thus, the xanthine exocyclic O2 interacts with two hydrogen bond donating waters.

*Crystal structures of E222Q bound to the purF operator and corepressors hypoxanthine, guanine and xanthine.*

To resolve the structural basis for the increased affinity for xanthine, yet unexpected decrease in binding for the physiological corepressors hypoxanthine and guanine of the isosteric E222Q mutant, we determined the crystal structures of the E222Q-hypoxanthine-*purF* operator complex, E222Q-guanine-*purF* operator complex, and E222Q-xanthine-*purF* operator complex. Superimposition of the C $\alpha$  carbons of these structures and the wild-type PurR-hypoxanthine-*purF* operator structure results in rmsds of 0.22 Å for the E222A-hypoxanthine-*purF* operator complex, 0.17 Å for the E222A-guanine-*pur* operator complex, and 0.23 Å for the E222A-xanthine-*purF* operator complex, all of which are within error of the coordinates. As in the wild-type PurR-DNA structure, the DNA is kinked about the central CpG base pair step. The DNA bending angle was calculated with CURVES (Lavery and Sklenar, 1988), excluding the first and last base pairs of the 16 base pair operator. The calculated bending angles, which are essentially identical to that calculated for the PurR-hypoxanthine-*purF* operator structure (Schumacher et al., 1994a), are 52° for the E222Q-hypoxanthine-*purF* operator structure, 49° for the E222Q-guanine -*purF* structure and 51° for the E222Q-xanthine-*purF*

operator structure. All other DNA properties were also nearly identical to those found in the wild-type PurR-corepressor-DNA structure (Schumacher et al., 1994a).

The replacement of the wild-type glutamate 222 with a glutamine is an isosteric replacement and was envisioned to test the loss of charge at this position. Furthermore, this substitution was expected to expand the corepressor binding specificity to include xanthine because of the dual function of its side chain with respect to hydrogen bonding, i.e., the O $\epsilon$  and N $\epsilon$  are acceptors and donors, respectively. Thus, the glutamine side chain would provide a hydrogen bond donor to the specificity determining water molecule, regardless of the exocyclic 2 position moiety of the corepressor. The glutamate to glutamine substitution would be predicted to result in a decreased affinity of the mutant for hypoxanthine and guanine due to loss of electrostatic complementarity between the electropositive N2 or C2 and a significant increase in binding of xanthine, which is electropositive (Figure 4.2). As predicted, affinity for xanthine was increased in this mutant. However, the equilibrium dissociation constants for hypoxanthine and guanine binding were significantly reduced compared to wild-type and were, in fact, similar to those obtained for the alanine mutant. The basis for the decrease in binding of the wild-type corepressors, hypoxanthine and guanine, was revealed in the structures of the E222Q-hypoxanthine-*purF* operator and E222Q-guanine-*purF* operator. These structures show that, unexpectedly, the Gln222 side chain rotates from the wild-type Glu222 location and now points out of the corepressor binding pocket. In this position, the Gln222 side chain no longer contacts the specificity determining water molecule. As

a result, the water molecule that is found tightly bound in the PurR-guanine-*purF* operator structure is very weakly anchored in the E222Q structures.

The relocation of the Glu222 side chain might occur because it allows it to function as an N-cap residue for helix 11 (Richardson and Richardson, 1988). Specifically, the Gln222 N $\epsilon$  hydrogen bonds to the peptide backbone carbonyl oxygen of Asp220 (Figure 4.3B). Another factor likely contributing to the relocation of this side chain is the concentrated negative charge found in the vicinity of the N-terminus of helix 11. This concentration of charge is predominantly forged from the side chains of Asp220 and Glu224. In the wild-type protein, this dense negative charge might function to repel the Glu222 side chain away from the helix N-terminus and thus, into the corepressor binding pocket. In contrast, the glutamine side chain is not repelled and, indeed, finds a more favorable role in helix N-capping. The relocation of the Gln222 side chain would also be predicted to relax specificity at the 2 position of the corepressor and thus allow xanthine binding. Indeed, in the structure of the E222Q-xanthine-*purF* operator complex this "removed side chain location" is maintained.

*Crystal structures of E222K bound to the purF operator and corepressors hypoxanthine, guanine or xanthine.*

To elucidate the stereochemical and structural basis by which E222K binds the natural corepressors, hypoxanthine and guanine with affinity similar to wild-type, and binds xanthine with affinity greater than wild-type, we determined three crystal structures: the E222K-hypoxanthine-*purF* operator complex, the E222K-guanine-*purF*

operator complex, and the E222K-xanthine-*purF* operator complex. Again as anticipated, the corepressor binding pockets of each of the E222K-corepressor-*purF* operator complexes reveals that none of the remaining wild-type protein-purine interactions are missing nor do the bases take radically different binding positions (Figure 4.3C). Superimposition of the C $\alpha$ s of each of these structures onto the corresponding atoms of the wild-type PurR-hypoxanthine-*purF* operator structure (1QPZ) confirms their structural identity and yields rmsds of 0.27 Å for the E222K-guanine-*purF* operator complex, 0.25 Å for the E222K-hypoxanthine-*purF* operator complex, and 0.24 Å for the E222K-xanthine-*purF* operator complex (all within the error of the coordinates). As observed in the wild-type PurR-hypoxanthine-*purF* structure, the DNA is kinked about the central CpG base pair step. The DNA bending angles of the *purF* site, which were calculated with CURVES (Lavery and Sklenar, 1988) and excluded the first and last base pairs of the 16 base pair operator because of their poor electron density, are 52°, 48°, and 47° for E222K bound to guanine, hypoxanthine, and xanthine, respectively, indicating that the DNA structure is essentially identical to that seen in the PurR-hypoxanthine-*purF* operator structure (Schumacher et al., 1994a). All other DNA properties were also nearly identical to those found in the wild-type protein-corepressor-DNA structure.

An overlay of the three E222K-purine-DNA structures shows that the xanthine ring is rotated slightly as compared to hypoxanthine and guanine. This is likely due to the change of the purine N3 from a hydrogen bond acceptor (hypoxanthine and guanine) to a hydrogen bond donor (xanthine). The rotation allows hydrogen bonding between xanthine N3 and the Asp275 carboxylate. In addition, the Arg190 side chain is

positioned slightly further from xanthine. This is likely caused by electrostatic repulsion between the Lys222 side chain and the Arg190 side chain.

To design a PurR with expanded, high affinity corepressor specificity, we predicted that a lysine side chain, having positive charge and hydrogen bond donors, would improve binding of xanthine to the E222K while reducing affinity for guanine and perhaps hypoxanthine. The affinity of E222K for the natural corepressors guanine and hypoxanthine is indeed reduced and is similar to that of the E222A mutant. This similarity may be explained by the conformation of the lysine side chain, which is oriented such that its positively charged side chain ammonium group is positioned away from corepressor, as well as by its relative disorder. The ~2 fold lower affinity of E222K for the natural corepressors, as compared to E222A, is likely due to poorer electrostatic complementarity due to the proximity of the positive charge borne by the lysine side chain, as all experiments were carried out at pH 7.5.

Also as predicted, the E222K mutant binds xanthine with higher affinity (five-fold) than wild-type PurR. The positive charge of lysine improves the electrostatic complementarity of the corepressor binding pocket to the more electronegative O2 of xanthine. In addition, a network of hydrogen bonds, mediated by water molecule W1, is observed. This network is qualitatively similar to that seen for wild-type PurR binding to guanine. However, in the E222K-xanthine-*purF* operator structure, the O2 of xanthine and the carbonyl oxygen of Phe221 both accept hydrogen bonds from water W1. W1 thus has the ability to accept hydrogen bonds from two donors, one of which is the amino group of Lys222. The participation of Lys222 in this network brings it closer to the

corepressor, improving electrostatic complementarity, while reducing the mobility of the lysine side chain, which unfavorably decreases its entropy. This unfavorable reduction of entropy must be compensated by the favorable enthalpy of hydrogen bond formation to the water molecule.

## Discussion

In the X-ray structure of the wild-type PurR-guanine-*purF* operator complex, residue Glu222 was recognized to be a primary determinant of corepressor specificity by virtue of a water mediated hydrogen bond with the guanine N2 and by the Glu222 contribution to the negative electrostatic environment of the binding pocket about guanine N2 (Schumacher et al., 1997). These interactions give rise to preference for the atom at the exocyclic 2 position of the purine to be a hydrogen bond donor, and thereby favor guanine binding over hypoxanthine binding. In addition, electropositive atoms, such as N2, are favored at the 2 position, again favoring guanine binding over hypoxanthine. Together, these interactions effectively preclude binding of xanthine, which is electronegative about the 2 position of the purine and carries a hydrogen bond accepting O2. In this work we have replaced Glu222 with alanine, glutamine, or lysine in order to assess directly the function of this residue and the structural consequences of such substitutions. Measurements of repression and *in vitro* purine binding have confirmed the critical role of this residue in corepressor binding.

Replacement of Glu222 with alanine results in a 1.4 fold increase in affinity for xanthine and diminishes the binding affinity for guanine and hypoxanthine by 6.7 and 2.1 fold, respectively. The larger decrease in affinity for guanine reduces the preference for guanine over hypoxanthine from 6.9 fold to 2.1 fold. The higher equilibrium dissociation constants of E222A for guanine and hypoxanthine arise in part from the loss of the negative charge provided by the Glu222 side chain. The less electronegative environment about the 2 position of the purine corepressors both reduces affinity for



hypoxanthine and guanine, which are electropositive at the 2 position, and increases affinity for xanthine, which is electronegative at the 2 position. The larger decrease in affinity seen for guanine is likely due to the loss of the water-mediated interaction between the Glu222 side chain and the guanine N2. However, as guanine is more electropositive than hypoxanthine at the 2 position, reduction in electrostatic complementarity also may contribute. The crystal structures of E222A in complex with the *purF* operator site and each of these purines reveal the stereochemical bases of their binding. Apart from the missing E222 side chain and the inability to find as many water molecules surrounding the base at the resolution to which these structures have been solved, no major changes are observed in the E222A mutant structures. The water network between the base and the protein is largely conserved. The water molecule that bridges the exocyclic 2 position of the purine and the Phe221 backbone carboxyl oxygen is conserved in each of the E222A-purine-*purF* operator structures. In the wild-type structure, this water donates hydrogen bonds to the Phe221 carboxyl oxygen and to O $\epsilon$ 1 of the E222 side chain. Therefore, the other two interactions this water is allowed must be hydrogen bond accepting interactions. In the E222A mutant, this restriction is removed, and this water molecule is able to either donate a hydrogen bond to the exocyclic O2 of xanthine or accept a hydrogen bond from the exocyclic N2 of guanine, i.e., it is no longer discriminating.

The replacement of the glutamate at position 222 with glutamine was designed to permit this mutant protein to retain its capacity to bind tightly to guanine and hypoxanthine via side chain O $\epsilon$ -water-N2 hydrogen bonds while gaining the ability to

bind xanthine with high affinity by side chain N $\epsilon$ -water-O2 hydrogen bonds.

Surprisingly, this is not the case. The Gln222 mutant does not make the expected hydrogen bonds in the binding pocket, but rather is rotated such that it can no longer reach the specificity-conferring water (W1) in the pocket. Glutamine is electrostatically neutral, as was the alanine, so in conjunction with its inability to hydrogen bond, E222Q has affinities for purines that are remarkably similar to those observed with E222A. The slight improvement in xanthine binding and decrease in binding of the natural corepressors might be the simple result of replacement with a longer neutral side chain, which further decreases the electronegativity of the pocket.

The Glu222Lys mutation was designed to increase the affinity of xanthine for PurR by making the environment about the 2 position of the bound purine more electropositive and by providing a hydrogen bond donor to complement the hydrogen bond-accepting O2 of xanthine. Indeed, E222K shows a 5.2 fold increase in affinity for xanthine and 14 and 3.4 fold decreases in affinity for guanine and hypoxanthine, respectively. These changes result in a protein with nearly equal, physiologically relevant, affinities for guanine, hypoxanthine, and xanthine. The affinities of E222K for the natural corepressors, hypoxanthine and guanine are nearly equal to those of E222A. However, the E222K protein has a 3.8 fold higher affinity for xanthine than E222A. Examination of the crystal structures of the E222K-purine-DNA complexes reveals the chemical basis for these observations.

In the complexes with the natural corepressors, the lysine side chain is disordered and is oriented away from the corepressor-binding pocket. Thus, it is not surprising that

E222K binds to the natural corepressors with affinity very similar to that of E222A. However the affinity of E222A is 2 and 1.6 fold higher than E222K for guanine and hypoxanthine, respectively. These small effects are likely due to changes in the electrostatic environment of the corepressor-binding pocket. Specifically, there is a decrease in electronegativity of the corepressor-binding pocket near the 2 position of the bound purine for E222K as compared to E222A. Consistent with this, the change from E222A to E222K disfavors the binding of guanine, which carries a stronger partial positive charge on its N2H1 ( $\delta=0.43$ ) and N2H2 ( $\delta=0.43$ ) atoms, more than the binding of hypoxanthine, which carries a lower partial positive charge on its C2H ( $\delta=0.10$ ) atom. We note that a similar ~2 fold decrease is observed in affinity for hypoxanthine upon the change of Glu222 to Ala. This correlates with a similar decrease in electronegativity of the corepressor binding pocket near the 2 position of the bound purine. Improved electrostatic complementarity likely accounts for the 1.4 fold increase in affinity for xanthine of E222A. However, the affinity of E222K for xanthine increases 5.2 fold, and E222K thus binds xanthine 3.8 fold more tightly than E222A does. Again, examination of the E222K-purine-DNA structures provides an explanation for this additional increase in affinity. In the E222K-xanthine-DNA structure, the Lys222 side chain is more ordered, and rather than pointing away from corepressor, is involved in a water-mediated hydrogen bonding network (Lys222 N $\epsilon$ -water-xanthine O2), reminiscent of that seen for wild-type PurR (Glu222 O $\epsilon$ -water-guanine N2). Indeed, a similar larger change in affinity, 6.2 fold, is observed for guanine binding when Glu222 is changed to Ala.

The preceding discussion, of course, makes the assumption that the free energies of the corepressor-free conformations of wild-type PurR and the mutants are very similar. For example, the reduced affinity of E222K for the natural corepressors could be due, in part, to higher stability of the aporepressor form of E222K relative to the aporepressor form of E222A, possibly due to a favorable hydrogen bond or salt bridge formed by the lysine side chain when in the aporepressor form. When the Glu222 side chain is replaced by glutamine in the apo PurR (CBD) structure, little impact on the stability of the protein is predicted, as Glu222 is on a solvent-exposed loop in the wild-type CBD structure (Schumacher et al., 1995).

Ligand specificity has been expanded or changed in a number of cases (reviewed in (Clackson, 1998)). Purine specificity in PurR has been expanded to include adenine (Lu et al., 1998b) or xanthine (Chapter 4), each with a single amino acyl residue replacement. The *Giardia lamblia* guanine phosphoribosyltransferase has been engineered to use either hypoxanthine or guanine as a high affinity substrate by introducing three point-mutations (Munagala et al., 2000). The specificity of the retinoic acid receptor, RXR, was manipulated such that affinity for its natural corepressor, 9-*cis*-retinoic acid, was increased 280-fold over three synthetic hormones with one substitution, but decreased 40-fold relative to the three synthetic hormones with another point mutation (Peet et al., 1998).

In conclusion, PurR mutant proteins E222Q and E222A are repressors with reduced preference for guanine over hypoxanthine, while the E222K protein exhibits extended corepressor specificity, as it has gained the ability to bind xanthine with

physiologically relevant affinity. Changes in corepressor affinity are highly correlated with changes in the electrostatic environment about the 2 position of the corepressor and with water mediated contacts between exocyclic atoms at the 2 position and residue 222 side chains.

### **Acknowledgments**

All PurR E222 mutant plasmids and proteins were made by Fu Lu in the Zalkin laboratory. Fu Lu also performed the *in vivo* repression and *in vitro* DNA and purine binding assays. The crystal structures of each of the E222Q-purine-DNA complexes were solved by Maria Schumacher, and those of the E222K-purine-DNA complexes were solved by Dennis Arvidson. I determined the structures of the three E222A-purine-DNA complexes.

Table 4.1. Affinities of wild-type and mutant purine repressors for guanine, hypoxanthine and xanthine.

PurR	Kd ( $\mu$ M) <sup>a</sup>			Relative Affinity		
	Gua	Hyp	Xan	Kd <sub>Hyp</sub> /Kd <sub>Gua</sub>	Kd <sub>Xan</sub> /Kd <sub>Gua</sub>	Kd <sub>Xan</sub> /Kd <sub>hyp</sub>
WT	1.6 +/- 0.5	11.0 +/- 1.6	233 +/- 58	6.9	146	21
E222A	10.7 +/- 1.7	23.0 +/- 1.8	174 +/- 26	2.1	16.3	7.6
E222Q	14.0 +/- 4.0	32.0 +/- 3.6	126 +/- 27	2.3	9.0	3.9
E222K	22.0 +/- 1.5	37 +/- 7.0	45.0 +/- 2.7	1.7	2.0	1.2

Table 4.2. Summary of selected crystallographic data collection and refinement statistics for E222A, Q, and K mutants bound to hypoxanthine (hyp), guanine (gua), or xanthine (xan) and *purF* operator.

<sup>a</sup>  $R_{\text{Symm}} = \Sigma |I_o - \langle I \rangle| / I_o$ , where  $I_o$  is the observed intensity and  $\langle I \rangle$  is the average intensity obtained from multiple observations of symmetry-related reflections.

<sup>b</sup>  $R_{\text{factor}} = \Sigma | |F_{\text{obs}}| - |F_{\text{calc}}| | / |F_{\text{obs}}|$ .

<sup>c</sup> Rmsds are the root mean squared deviations of the bond lengths and bond angles are from their respective ideal values as implemented in TNT (Tronrud *et al.*, 1987).

<sup>d</sup> The Ramachandran analysis was generated with PROCHECK (Laskowski *et al.*, 1993).

## Data Collection

Mutant-corepressor  
Cell Constants

	E222A-hyp	E222A-gua	E222A-xan	E222Q-hyp	E222Q-gua	E222Q-xan	E222K-hyp	E222K-gua	E222K-xan
a (Å)	C222 <sub>1</sub> 176.16	C222 <sub>1</sub> 175.68	C222 <sub>1</sub> 176.04	C222 <sub>1</sub> 176.02	C222 <sub>1</sub> 176.07 Å	C222 <sub>1</sub> 175.82	C222 <sub>1</sub> 175.77	C222 <sub>1</sub> 175.79	C222 <sub>1</sub> 176.56
b (Å)	94.47	94.76	95.46	95.00	95.15 Å	94.66	94.99	94.84	95.60
c (Å)	81.63	81.50	81.92	81.52	81.56 Å	81.07	81.86	81.77	81.60
Unique reflections (#)	16,715	17,564	16,117	26,982	34,743	14,262	13,404	15,853	17,197
Total Reflections (#)	24,552	67,088	18,025	49,927	56,742	67,253	36,916	40,109	52,223
Completeness (%)	81	93	88	92	96	91	92.2	83.5	99
R <sub>s</sub> ym (%) <sup>a</sup>	9.9	10.9	8.8	8.3	6.6	7.4	9.53	8.2	8.6
I/σ(I)	7.8	9.0	10.5	6.0	6.2	6.7	10	7	5.9

## Refinement Statistics

Resolution (Å)	2.70	2.71	2.76	2.7	2.8	2.9	2.9	2.7	2.8
R <sub>F</sub> actor (%) <sup>b</sup>	18.3	19.2	18.8	18.5	16.9	18.1	19.1	22.0	21.3
Number of atoms	3050	3069	3032	3007	3008	3008	3023	3023	3043
Number of solvent molecules	35	39	23	53	39	25	11	10	30

## Stereochemistry

R <sub>ms</sub> d <sup>c</sup>									
Bond Lengths (Å)	0.014	0.015	0.013	0.014	0.014	0.017	0.007	0.006	0.009
Bond Angles (°)	1.581	1.690	1.544	2.020	1.745	1.810	0.61	0.68	0.75
B Correlations (Å <sup>2</sup> )	4.11	4.00	3.51	4.37	3.77	3.19	3.79	4.27	3.87

## Ramachandran Plot Analysis<sup>d</sup>

Region

most favored (#/%)	247/82.9	252/84.3	257/86.2	254/85.2	258/86.9	250/83.9	250/83.9	238/79.9	244/81.9
additional allowed	46/15.4	42/14.4	38/12.8	41/13.8	36/12/1	44/14.8	44/14.8	56/18.8	51/17.1
generously allowed	3/1.0	3/1.0	1/0.3	1/0.3	1/0.3	2/0.7	2/0.7	2/0.7	0/0
disallowed	2/0.7	2/0.7	2/0.7	2/0.7	2/0.7	2/0.7	2/0.7	2/0.7	3/1.0



Figure 4.1. A close-up view of the corepressor binding pockets of the wild-type PurR-hypoxanthine-*purF* operator complex, in A, and PurR-guanine-*purF* operator complex, in B.

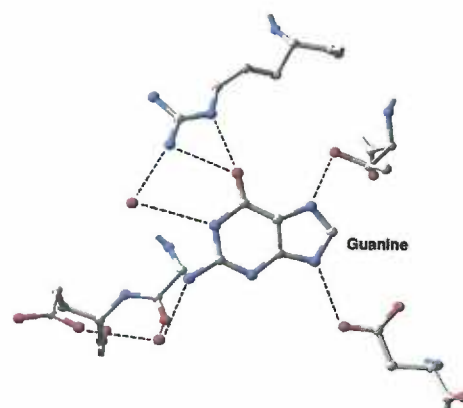
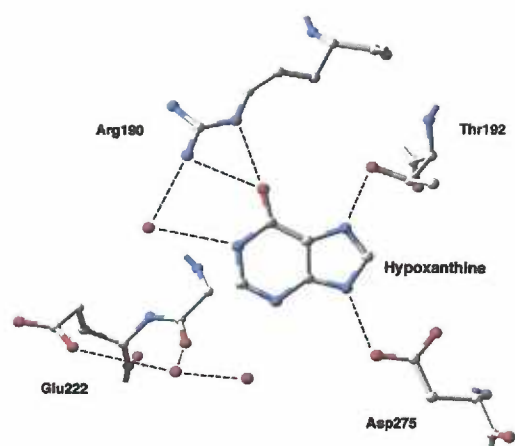
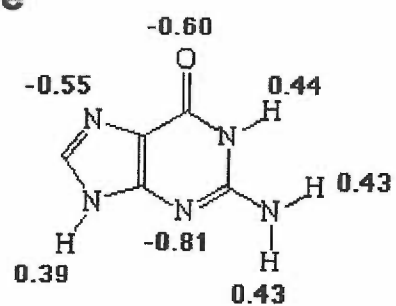
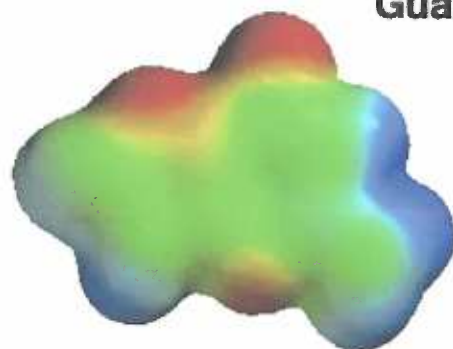
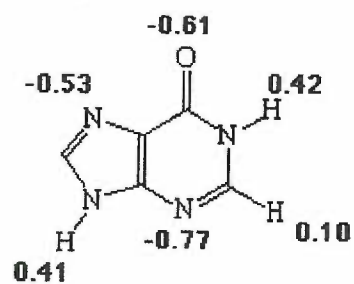
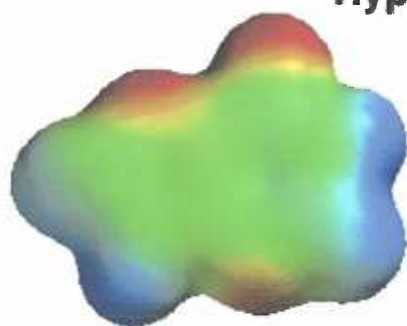


Figure 4.2. The electrostatic surfaces and partial charges of: A. guanine, B. hypoxanthine, and C. xanthine, as calculated with SPARTAN software by Dr. Arthur Glasfeld.

**Guanine**



**Hypoxanthine**



**Xanthine**

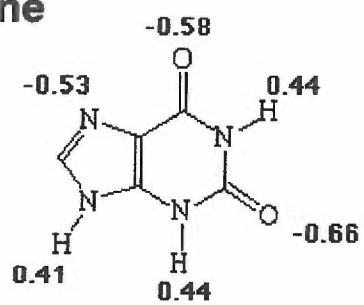
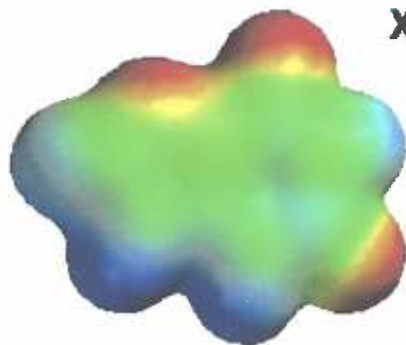
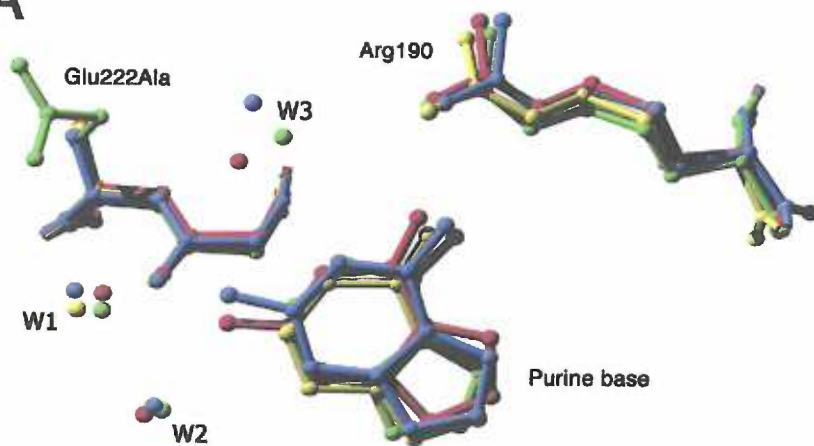
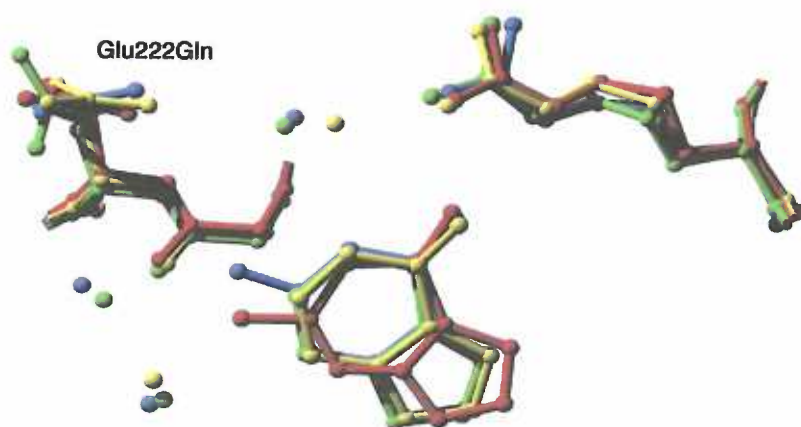


Figure 4.3. Views of the corepressor binding pocket of the mutants after the superimposition of their C $\alpha$  carbon atoms onto those of the wild-type PurR protein bound to hypoxanthine. Water molecules referred to in the text are marked W1, W2, or W3. The key specificity-determining water is W1. A. The E222A binding pocket structures with guanine (blue), hypoxanthine (yellow), xanthine (red), and PurR-hypoxanthine (green). B. The E222Q binding pocket structures showing complexes with purines colored as in A. C. The E222K binding pocket structures, colored as in A. With the exception of those made by E222, all other wild-type protein-purine contacts are found in each E222A mutant-purine complex.

**A**



**B**



**C**

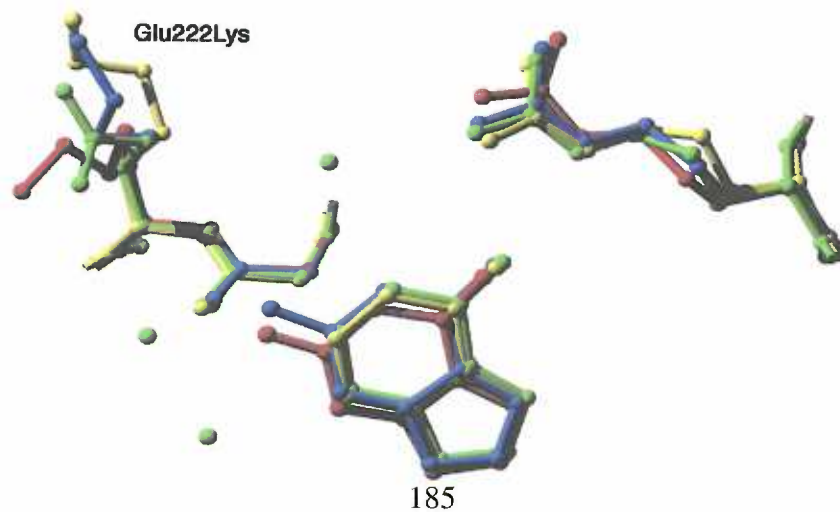
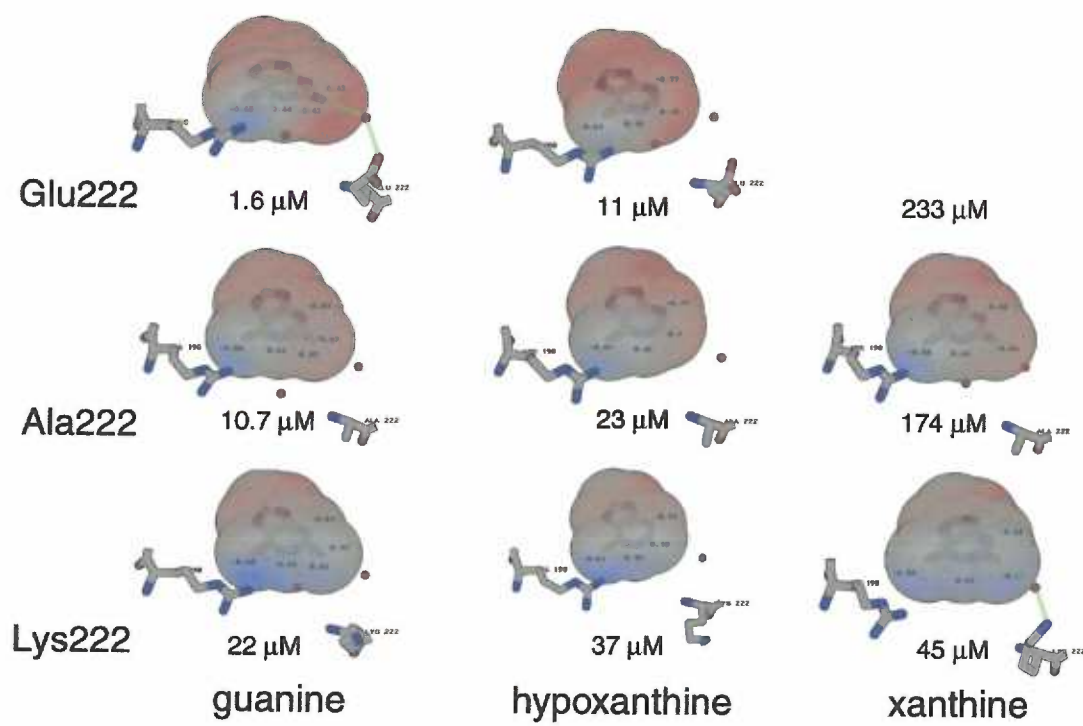


Figure 4.4. Electrostatic environment of corepressors. The electrostatic potential for each wild-type, E222A, and E222K protein structure was calculated and interpolated to the accessible surface for each corepressor. Red surfaces are more electronegative and blue surfaces are more electropositive.





**Chapter 5. Towards redesign of corepressor specificity of the *Escherichia coli* purine repressor by substitution of residues 190 and 192**

Joy L. Huffman and Richard G. Brennan

*Dept. of Biochemistry and Molecular Biology, Oregon Health Sciences University*

## Abstract

The *Escherichia coli* purine repressor (PurR) represses expression of *pur* regulon genes when it has bound a corepressor, guanine or hypoxanthine. The exquisite specificity of corepressor binding is brought about by multiple side chain-purine base contacts. Surprisingly, another member of the LacI/GalR family, the ribose repressor RbsR, has nearly the same complement of side chains, yet binds ribose. PurR and RbsR diverged most recently among LacI/GalR family members on the phylogenetic tree. These proteins are 46% identical, with all but two of the key residues in the ligand binding pocket conserved. Specifically, an arginine at position 190 in PurR corresponds to a lysine in RbsR, and the threonine at position 192 is replaced by a proline in RbsR. Previously, structure-based studies proved the importance of Arg190<sup>1</sup> in determining 6-oxopurine preference and 6-aminopurine or 6-methylpurine exclusion. We hypothesized that changing the two residues that differ between PurR and RbsR might change the ligand specificity from purines to ribose.

Replacement of Arg190 with Lys resulted in a functional purine repressor in which binding of hypoxanthine was weakened slightly and affinity for ribose was increased. Replacement of Thr192 with Pro resulted in a sharply destabilized repressor that no longer bound hypoxanthine or ribose. The R190K/T192P double mutant displayed no affinity for hypoxanthine, but ribose binding was somewhat increased over that of the T192P alone. The crystal structure was determined for the ternary complex of R190K with hypoxanthine and *purF* operator DNA. The change in affinity observed for hypoxanthine correlates well with the loss of a hydrogen bond between the 6 position of the purine ring and the residue at position 190. In conclusion, the replacement of PurR residues Arg190

and Thr192 with the analogously positioned RbsR residues, Lys and Pro, does not allow PurR to use ribose as its small molecule corepressor.

<sup>1</sup> For clarity, one-letter amino acid codes will be used in naming mutant proteins, as in R190K. Three-letter codes will refer to individual residues, such as Arg190. Thus, Arg190Lys will refer to the lysine residue that has replaced an arginine at position 190, whilst R190K will be the protein containing this substitution.

## Introduction

The *Escherichia coli* purine repressor (PurR) binds to a conserved 16 base pair operator sequence and regulates transcription of multiple *pur* regulon genes, including nine operons for *de novo* purine nucleotide biosynthesis, the *purR* gene and other genes with functions related to nucleotide metabolism (Zalkin and Nygaard, 1996). Binding of a corepressor, hypoxanthine or guanine, induces a conformational change in PurR that allows the signal of this binding to be passed to the DNA binding domain and, consequently, cognate DNA to be bound (Rolfes and Zalkin, 1990b). PurR is a 341 amino acid residue polypeptide that functions as a dimer. Each subunit is composed of an N-terminal DNA binding domain (residues 1-60) and a larger C-terminal corepressor binding and dimerization domain (residues 61-341), termed the CBD. The crystal structures of free CBD and intact PurR in complex with hypoxanthine and cognate DNA have been solved (Schumacher et al., 1995; Schumacher et al., 1994a).

Inspection of the corepressor binding pocket of PurR bound to hypoxanthine implicated residue Arg190 as a key specificity determinant of PurR for 6-oxopurines and made clear the preference of PurR for guanine and hypoxanthine over adenine (Lu et al., 1998a). Specifically, the guanidinium group of Arg190 donates two hydrogen bonds to the exocyclic O6 of hypoxanthine or guanine, thus discriminating against adenine, which has an exocyclic amino group at position 6. This amino group can only donate hydrogen bonds, as can the Arg190 side chain, and therefore adenine is disfavored due to the close proximity of the side chain and purine 6 position and its inability to accept hydrogen bonds. In addition, the Arg190 guanidinium group contributes to the electropositive environment about the 6 position of corepressor, thus providing additional corepressor

selectivity for the more electronegative O6 of guanine and hypoxanthine and discrimination against the more electropositive N6 of adenine.

Previously, biochemical and structural studies on two site-directed mutants at position 190 elucidated the requirement for 6-oxopurines (Lu et al., 1998b). When Arg190 was changed to Ala, corepressor specificity was expanded to include adenine and 6-methylpurine, with similar albeit slightly higher  $K_d$ s observed for all of the corepressors. The structures of each of the ternary complexes revealed a rotation of the Ser124 hydroxyl group with respect to its position in wild-type PurR. Moreover, the polypeptide backbone around Ser124 moves to relieve the strain found in the wild-type structure, where Ser124 is always a Ramachandran outlier. In the R190A structure, Ser124 makes a compensatory hydrogen bond with the exocyclic 6 positions of each of the bases, whereas in wild-type PurR this residue does not contact corepressor directly. When Arg190 was changed to glutamine, wild-type affinities for hypoxanthine and guanine were retained, and a 75-fold higher affinity for adenine—to the physiologically relevant level of 33  $\mu$ M—was achieved by new direct and water-mediated contacts.

In light of these results, we hypothesized that even more dramatic changes in corepressor specificity might be feasible. To test this hypothesis, we studied the sequence alignment of the LacI/GalR family and noted the high sequence homology between PurR and RbsR, the ribose repressor. PurR and RbsR diverged from each other the most recently on the phylogenetic tree (Figure 5.1A, (Nguyen and Saier, 1995). While PurR represses transcription by binding cognate DNA when it has bound its small molecule ligand, RbsR induces transcription by dissociating from cognate DNA when it has bound its small molecule ligand. The two proteins are 46% identical and differ at only two of the

residues that directly contact corepressor in the PurR holo complex structure (Figure 5.1B). These two differing residues are Arg190, which is a lysine in RbsR, and Thr192, which is a proline in RbsR. Unfortunately, the crystal structure of RbsR is unknown. However, the structure of another closely related protein, the Ribose Binding Protein (RBP), has been solved. Many of the ligand-binding residues in this protein are also conserved with those of PurR and, putatively, RbsR (Figure 5.2). Thus, these proteins use nearly identical residues to bind a completely different ligand. Interestingly, both PurR and RBP use arginine and threonine to bind their respective ligands, but in RBP the arginine comes from a different strand from that of the threonine. Regardless, their three-dimensional locations are the same.

In this chapter, the structure-based design and analysis of three mutant PurR proteins, R190K, T192P, and R190K/T192P, are described. To study the function of the mutations *in vitro*, we purified the mutated proteins to assess their corepressor and DNA binding affinities. A method for determining relative affinities of wild-type and mutant PurR proteins for ribose is described. Finally, we describe the crystal structure of the R190K-hypoxanthine-*purF* operator complex. Replacement of Arg190 by lysine reduced the preference of PurR for hypoxanthine, while increasing the affinity for ribose. The structure of R190K bound to hypoxanthine reveals the structural and chemical features contributing to lowered hypoxanthine affinity. However, neither the T192P nor R190K/T192P PurR mutants crystallized in complex with hypoxanthine and the *purF* operator. Furthermore, none of the mutants crystallized in the presence of ribose and *purF* operator. A model of the T192P complex will be presented, as will discussion of possible reasons for our inability to switch corepressor specificity or even increase ribose

affinity by substitution of these residues.

## Materials and Methods

### *Site-directed mutagenesis.*

Site-directed mutagenesis by the procedure of Kunkel *et al.* (Kunkel, 1985) was carried out as described previously (Choi et al., 1994). Single-stranded *purR* template DNA was prepared from plasmid pPR1005-2 (Rolfes and Zalkin, 1990b). The double stranded template pPR1005-2, containing the wild-type *purR* gene, was transformed into *E. coli dut<sup>-</sup> ung<sup>-</sup>* strain CJ236, which was subsequently infected with M13K07 helper phage. This strain produces uracil-containing DNA when replicated. The phagemids were purified and the uracil-containing single-stranded DNA extracted. Mutagenic oligonucleotide primers were annealed to this single-stranded DNA and the complementary strand was synthesized using T4 DNA polymerase. The remaining nick in the DNA at the 5' end of the oligonucleotide was ligated to the newly synthesized strand using T4 DNA ligase to form covalently closed circular DNA. This DNA was then transformed into MV1190 cells, which have a functional uracil N-glycosylase, such that the mutagenized strand would be replicated and the uracil-containing template strand would be inactivated. Colonies were picked, and the DNA was extracted and sequenced.

### *Overexpression and purification of PurR.*

Wild type and mutant *purR* genes in pPR1005-2 were transferred via flanking *NdeI* and *HindIII* sites to pET24(a) for overexpression. Overexpressed proteins were purified as described (Choi and Zalkin, 1992). The elution profiles of PurR from DEAE Sepharose and heparin agarose columns were identical for wild type and mutants. For purified PurR,



the protein concentration was determined spectrophotometrically at 280 nm using  $A_{280}$  (1%) of 12.9 (Rolfes and Zalkin, 1990b).

*PurR equilibrium binding to corepressor and operator DNA.*

Fluorescence polarization was used to determine the affinities of wild-type and mutant PurR proteins for hypoxanthine, ribose, and *purF* operator under a number of conditions. Experiments were done using a PanVera Beacon Fluorescence Polarization System (PanVera Corporation). Samples were excited at 490 nm and emission was measured at 530 nm. A 5' fluoresceinated oligonucleotide corresponding to the *E. coli purF* operator (Oligos, Etc.) (5'-F-AAAGAAAACGTTTGCGTACCCCTACGCAAACGTTTCTTT-3') was self-annealed by heating to 80 °C, followed by flash cooling, to form a stem-loop structure with the *purF* operator motif at the center of the base-paired region. This oligonucleotide is referred to as the "F-loop", for fluoresceinated stem-loop *purF* operator. Binding was assayed in a 1 ml volume at 25 or 35 °C. The components of "corepressor dependent" binding experiments were: 2 nM F-loop DNA, 250 mM potassium glutamate, 150 mM NaCl, 10 mM magnesium acetate, 1 mM EDTA, 100 mM potassium HEPES, pH 7.5, and 1.0 µg/ml poly d(IC) (Rolfes and Zalkin, 1990b). The components of "corepressor independent" experiments were: 2 nM F-loop DNA, 10 mM potassium HEPES, pH 8.0, 100 mM KCl, 1 mM EDTA, and 1.0 µg/ml poly d(IC) (*ibid.*).

To determine the corepressor-dependent binding of protein to *purF* operator, saturating amounts (200 µM) of hypoxanthine or 2.0 M ribose were included in the cuvette and protein was titrated. In one experiment with wild-type PurR, no corepressor

was added to the cuvette, and no protein:DNA binding was detected, thus confirming the efficacy of the buffer system. The PurR protein was diluted serially between 100  $\mu\text{M}$  and 1.0  $\mu\text{M}$  and titrated into the cuvette. The same type of assay was used to measure the binding of hypoxanthine to protein indirectly, as PurR does not appreciably bind DNA in the absence of corepressor. It is assumed that as protein binds corepressor, it will immediately bind DNA, and thus the change in polarization can be measured as an indirect function of corepressor binding. In these hypoxanthine-binding experiments, saturating amounts of protein (100 nM) were included in the cuvette, and hypoxanthine was titrated. After each addition of either protein or corepressor, samples were incubated in the Beacon instrument for 30 s before a measurement was taken. This 30 s incubation allowed equilibrium to be reached, as longer incubations showed no change in polarization (data not shown). The mP (milliPolarization) at each titration point represents the average of eight measurements integrated over 6 s.

The data of each binding isotherm were analyzed by least squares regression using SigmaPlot software (Jandel Corporation). The following equation was applied:

$$A = \left\{ \frac{(A_{\text{bound}} - A_{\text{free}})[P]}{K_d + [P]} \right\} + A_{\text{free}}$$

where A is the polarization measured at a given total concentration of peptide, [P],  $A_{\text{free}}$  is the initial polarization of the free DNA, and  $A_{\text{bound}}$  is the polarization of maximally bound DNA (Lundblad et al., 1996). Least squares regression was used to determine  $A_{\text{free}}$ ,  $A_{\text{bound}}$ , and  $K_d$ .

RbsR binds ribose with  $K_d \approx 10 \text{ mM}$ , so we predicted that the  $K_d$  of wild-type PurR might be several orders of magnitude higher. Using fluorescence polarization, ribose

cannot be titrated to the concentrations required to generate an equilibrium binding isotherm in a manner similar to that of hypoxanthine due to the large change in viscosity upon addition of 2 M ribose. Therefore, the binding assay was modified to determine approximate relative affinities of wild-type and mutant PurR proteins for ribose. All PurR:ribose binding experiments were done at 35 °C. This temperature decreases the viscosity of the solution significantly and does not have a large effect on  $K_d$  for hypoxanthine or F-loop DNA. Individual cuvettes containing increasing amounts of ribose are treated as separate experiments that each correspond to one data point, which are combined to generate binding curves. In one cuvette, the polarization of F-loop DNA, corepressor-dependent binding buffer, and 0 to 3 M ribose is first measured to obtain  $A_0$ , the value for free DNA. 100 nM PurR is then added and the polarization again measured to obtain  $A_1$ . An increase in polarization indicates PurR binding to F-loop DNA, which is assumed to be the result of PurR using ribose as corepressor. Finally, 200  $\mu$ M hypoxanthine is added to the cuvette, and the polarization measured is  $A_{max}$ . If  $A_{max}$  is greater than  $A_1$ , then hypoxanthine is able to function further as a corepressor and ribose binding to PurR is not saturating. The equation, percent binding (%) =  $(A_1 - A_0)/(A_{max} - A_0)$ , is used to determine the percent of PurR:DNA binding induced by ribose. Data are plotted as percent binding versus ribose concentration. Using this method, binding isotherms for wild-type and R190K have been obtained. Relative affinities are derived from comparison of the ribose concentrations at which the DNA is 50% bound. Because mutants R190K/T192P and T192P did not bind hypoxanthine, maximum DNA binding could not be calculated for these proteins. The curves presented for these mutants are

plotted as the change in mP upon protein addition ( $A_1 - A_0$ ) against the concentration of ribose.

#### *Crystallization and data collection.*

The PurR mutant proteins were stored at concentrations of between 8 and 10 mg/mL in 200 mM potassium phosphate, pH 7.5, 0.1 mM EDTA, 0.2 mM DTT, 5% glycerol. For crystallization, the 16 bp *purF* operator with a 5' overhanging T (Oligos, Etc.), 5' - TACGCAAACGTTTTCTT - 3' (Schumacher et al., 1994b), was annealed to its complement with a 5' overhanging A and buffered in 10 mM sodium cacodylate, pH 6.5. The hypoxanthine stock solution was 100 mM in 110 mM KOH. Crystals of R190K-hypoxanthine-*purF* operator were grown using the hanging drop vapor diffusion method, as described previously (Schumacher et al., 1994b). Briefly, 100  $\mu$ M protein solution (8 mg/mL), saturated with hypoxanthine corepressor, was mixed with *purF* operator so that a 1:1.5 molar ratio of protein dimer:duplex DNA solution resulted. This solution was then mixed 1:1 with the reservoir solution, which is 25% PEG 4000 (BDH), 0.4 M ammonium sulfate, 50 mM cobalt hexammine chloride, and 0.1 M ammonium phosphate, pH 7.5. Crystallization is carried out at  $\sim 22^\circ\text{C}$ . Crystals first appeared as thin plates at the interface of the phase separation that occurs in the hanging drop. After several weeks to months, these plates yield large (0.6 mm x 0.3 mm x 0.2 mm) rhombohedrons. The resulting crystals are isomorphous to wild-type PurR-hypoxanthine-*purF* operator crystals. X-ray intensity data were collected at  $\sim 22^\circ\text{C}$  on an ADSC area detector using a Rigaku RU-200 X-ray generator.

### *Structure determination and refinement.*

The R190K-hypoxanthine-*purF* operator structure was solved using difference Fourier techniques utilizing the wild-type PurR-hypoxanthine-*purF* operator structure (1PNR) as the starting model after removing the hypoxanthine, side chain at position R190, and all solvent. The asymmetric unit contains one PurR monomer, one purine base, and a *purF* operator half-site. Rigid body refinement, followed by XYZ (positional) refinement, and finally XYZ and B (thermal) refinement, as implemented in TNT (Tronrud et al., 1987), was carried out until convergence of the R-factor and stereochemical parameters. Hypoxanthine, the R190K side chain, and solvent molecules were built into the structures in the final stages of refinement using the molecular graphics program O (Jones, 1991).  $F_o - F_c$  electron density difference maps were calculated to ensure the unbiased placement of side chains, purines, and solvent molecules. The components in question were removed from the model, which was then refined 20 cycles before inspection of the omit density map.

A molecular model of the T192P -hypoxanthine-*purF* operator complex was obtained by simply mutating the residues in question in the wild-type structure using O. Reasonable hydrogen bonding was enforced, but the rest of the protein was not moved in the mutant model structure.

## Results

### *Corepressor-dependent binding of wild-type and mutant PurR proteins to purF operator DNA.*

Fluorescence polarization was used to determine binding affinities of wild-type and mutant proteins to *purF* operator as induced by hypoxanthine or ribose. Binding constants are summarized in Table 5.1 and representative equilibrium binding isotherms are presented in Figure 5.3. At 35 °C, the following  $K_d$  values were obtained: wild-type PurR, 4.7 +/- 0.4 nM; R190K, 9.3 +/- 1.4 nM; R190K/T192P, no specific binding; T192P, 1.59 +/- 0.10  $\mu$ M. Wild-type values are in close agreement with previous data (Choi et al., 1994; Glasfeld et al., 1999). Experiments carried out at 25 °C yielded similar values for wild type (2.5 +/- 0.4 nM) and R190K (7.9 +/- 0.6 nM), and no binding was detected with T192P or R190K/T192P (data not shown). In the presence of 2 M ribose rather than hypoxanthine, similar affinities were observed for wild-type (4.7 +/- 1.7 nM) and R190K (8.9 +/- 4.6 nM). In the absence of hypoxanthine or ribose, no protein:DNA binding was detected. Thus, both wild-type and R190K proteins are induced to bind *purF* operator DNA by 2 M ribose or micromolar concentrations of hypoxanthine, while R190K/T192P and T192P proteins do not bind the *purF* operator in the presence of either small molecule.

### *Equilibrium binding of hypoxanthine to wild-type and mutant proteins.*

To evaluate the proposed role of Arg190 for purine corepressor specificity, we measured binding affinities of hypoxanthine for wild-type PurR and mutants R190K, R190K/T192P, and T192P. The resulting titration curves are shown in Figure 5.4. At 35

°C, wild-type PurR binds hypoxanthine with an equilibrium dissociation constant of  $2.2 \pm 0.2 \mu\text{M}$ . In the R190K repressor, the binding affinity for hypoxanthine was somewhat decreased ( $K_d = 9.7 \pm 1.4 \mu\text{M}$ ), and neither R190K/T192P nor T192P bound hypoxanthine at concentrations of up to 1.5 mM. Binding of T192P to hypoxanthine could not be quantified as binding the assay is dependent upon specific, high affinity binding to *purF* operator.

*Corepressor-independent binding of purF operator DNA.*

The ability of wild-type and R190K/T192P mutant proteins to bind *purF* operator DNA regardless of corepressor binding was measured by employing a "corepressor independent" buffer system. This system has been shown to allow cognate DNA binding in the absence of corepressor (Rolfes and Zalkin, 1990b). Representative binding isotherms are shown in Figure 5.5. The equilibrium binding constant obtained for wild-type PurR binding to the *purF* operator DNA was  $1.5 \pm 0.3 \text{ nM}$ , that of R190K was  $3.6 \pm 0.5 \text{ nM}$ , that of R190K/T192P was  $93.0 \pm 11.6 \text{ nM}$ , and that of T192P was  $648 \pm 74 \text{ nM}$ . The sharp increase in  $K_d$  among the T192P-containing mutants may be attributed to lower specific activity of the proteins, but specific activity cannot be determined for these mutants as there is no control for binding.

*Binding of ribose to wild-type and mutant PurR proteins.*

The relative abilities of the wild-type and mutant proteins to use ribose as a corepressor were then examined. The apparent affinity of R190K for ribose is higher than that of wild-type PurR (Figure 5.6). The wild-type protein is half-saturated at 1.5 M

ribose, while the R190K protein is half-saturated at 1.0 M ribose. This difference is significant in that it is repeatable. The affinities of the other mutants can be qualitatively described by comparison of the concentration of ribose required to induce an increase of 10 mP upon addition of protein. Wild-type PurR requires 1.0 M ribose to induce this change in mP, while R190K/T192P requires 2.0 M ribose, and T192P requires 2.5 M ribose to induce a 10 mP increase. Based on these findings, we hypothesize that the Arg190Lys mutation contributes to an increase in ribose affinity, but that the Thr192Pro mutation is highly destabilizing to the protein. This effect is rescued to a degree by addition of the Arg190Lys substitution to the T192P mutant protein.

*Crystal structure of R190K bound to the purF operator and hypoxanthine.*

To determine the structural basis of R190K binding to the physiological corepressor hypoxanthine and its decreased affinity for hypoxanthine, we determined the crystal structure of the R190K-hypoxanthine-*purF* operator complex. Superimposition of the C $\alpha$  carbons of this structure and the wild type PurR-hypoxanthine-*purF* operator structure confirms that they are identical within error of the coordinates, with a root mean-squared deviation (rmsd) of 0.26 Å. As in the wild-type PurR-DNA structure, the DNA is kinked about the central CpG base pair step. The DNA bending angle was 52° as calculated with CURVES (Lavery and Sklenar, 1988), excluding the 5' overhanging nucleosides and the first and last base pairs of the 16 base pair operator because of their poor electron density. This value indicates that the DNA conformation is essentially identical to that of wild-type PurR-hypoxanthine-*purF* operator (Schumacher et al.,



*Model of the T192P-hypoxanthine-purF operator structure.*

The T192P mutation was qualitatively noted to be structurally destabilizing to PurR in that proteins carrying this mutation were more susceptible to proteolytic cleavage upon storage at -20 °C, being cleaved completely into respective corepressor and DNA binding domains within a week of purification, compared to wild-type, which can last months to years at -20 °C (data not shown). PurR can be specifically cleaved by trypsin or chymotrypsin in its linker region (residues 52-60) between the DNA and corepressor binding domains (Choi and Zalkin, 1994). PurR is protected from this cleavage when it binds cognate DNA in the presence of corepressor, due to a coil-to-helix transition in the linker region. A mutation that altered the overall fold of PurR, decreasing its ability to take the "DNA binding" conformation, would likely increase its susceptibility to proteolytic digestion. Multiple crystallization attempts of both the T192P and R190K/T192P-corepressor-DNA complexes were not successful. To obtain some insight into this failure, we modelled the threonine to proline mutation in the wild-type PurR structure.

Inspection of the model (Fig. 5.7), reveals that the loop on which Arg190 and Thr192 reside, residues 185-193, is in fact a tight turn at the N-terminal end of an  $\alpha$  helix. This loop has the sequence Pro-Gly-Pro-Leu-Glu-Arg<sup>190</sup>-Asn-Thr<sup>192</sup>-Gly, where Arg190 and Thr192 both contact corepressor. Changing Thr192 to a proline brings the number of prolines to three within the space of eight residues. The Thr192Pro mutation would likely have some effect on the shape of the turn that alters the position of Arg190, which has been shown to be key for binding 6-oxopurines (Lu et al., 1998b; Schumacher et al.,

1994a). This mutation may also disrupt the tight packing of the  $\beta$  sheet and  $\alpha$  helix before and after the turn, which could lead to the lower stability of the T192P protein.

Furthermore, the Thr192Pro C $\delta$  in the model is 3.3 Å from the Arg190 C $\beta$ . Replacing Thr192 with Pro in the R190K structure leaves the Thr192Pro C $\beta$  3.8 Å from the Arg190Lys C $\beta$ , a less energetically unfavorable distance for carbon-carbon van der Waals interactions. The nearest approach of the wild-type Thr192 and Arg190 side chains is 4.10 Å, between Thr192 O $\delta$  and Arg190 C $\beta$ .

## Discussion

The goals of these experiments were to 1) evolve the specificity of PurR from a purine corepressor binding protein to one that could bind ribose, a puckered sugar ring, using our understanding of ligand binding, and 2) explore the effects of the Arg190Lys mutation. Previous experiments showed that such manipulations might be possible (Lu et al., 1998b). By changing individual residues in the corepressor binding pocket, the affinity of PurR for corepressor has been expanded to include purines other than hypoxanthine and guanine. Effective removal of the Arg190 side chain by mutation to alanine resulted in an *in vitro* increase in affinity for 6-methylpurine and adenine, with a concomitant decrease in affinity for hypoxanthine and guanine such that all four purines were bound with nearly equal equilibrium dissociation constants. Substitution of Arg190 with glutamine allowed the repressor to retain high affinity binding of hypoxanthine and guanine and expanded the specificity of the repressor by allowing high affinity binding of adenine (~80-fold better than wild-type). In a different but related set of experiments, substitution of Glu222 with glutamine expanded the specificity of PurR to include xanthine (Chapter 4, this thesis).

Based on these findings we looked to other members of the LacI/GalR family for homologies that might lead to further "evolution" experiments. PurR and RbsR, the ribose repressor, are very similar, differing at only two of the residues that directly contact corepressor in the PurR holo complex structure. These two differing residues are Arg190, which is a lysine in RbsR, and T192, which is a proline in RbsR. Additionally, two residues that stack against corepressor, Tyr73 and Phe 74, are switched to the corresponding Phe73 and Tyr74 in RbsR.

In work described here, the affinities of R190K, T192P, and R190K/T192P for cognate DNA and, where possible, hypoxanthine and ribose were determined. To do so, a method for determining relative affinities of wild-type and mutant PurR proteins for ribose was developed. The crystal structure of the R190K PurR-hypoxanthine-*purF* operator complex was determined and a hypothetical model of the T192P mutant structure was constructed, as neither the T192P nor R190K/T192P PurR mutants crystallized in complex with hypoxanthine and the *purF* operator. Furthermore, none of the mutants crystallized in the presence of ribose and *purF* operator.

Replacement of Arg190 by lysine—a relatively conservative change—reduced the preference of PurR for hypoxanthine by ~4 fold, while increasing the affinity for ribose by ~1.7 fold, supporting the idea that this residue is not only important for conferring specificity for purines or sugar rings but that the smaller size and positive charge is necessary for ribose binding in RbsR. The crystal structure of R190K bound to hypoxanthine and the *purF* operator reveals that the moderate loss of affinity for hypoxanthine stems from the loss of the Arg190 Nε-hypoxanthine O6 hydrogen bond. Unfortunately, without the structure of the ribose-bound complex, we do not know the "real" basis for the increase in ribose affinity. The ribose binding protein (RBP) structure does not give any clues as to why a lysine works better, as this protein uses an arginine at this position to bind ribose (Mowbray and Cole, 1992).

Corepressor binding by the T192P mutant was not detected under any circumstances, and binding of *purF* operator was severely impaired ( $K_d = 1.6 \mu\text{M}$ ) under corepressor-dependent binding conditions. The R190K/T192P mutant protein bound *purF* operator with low but detectable affinity, and bound ribose better than the T192P mutant,

indicating that the Arg190Lys mutation slightly restores functionality to the T192P mutant, as determined by a qualitative assay, the results of which are shown in Figure 5.6. In a more quantitative assay, we determined the ability of each of these proteins to bind cognate DNA under "corepressor-independent" conditions that had been determined previously (Rolfes and Zalkin, 1990b), which allows high affinity DNA binding by PurR in the absence of corepressor. The wild-type and R190K proteins showed nearly equal, high affinity binding ( $K_d = 1.5$  and  $3.6$  nM, respectively). The R190K/T192P protein bound DNA with significantly lower affinity ( $K_d = 93$  nM), and the T192P mutant affinity was 6-7 fold lower than that of the double mutant ( $K_d = 649$  nM). The  $\Delta\Delta G$  for R190K binding to the *purF* operator, relative to wild-type PurR, was calculated to be  $0.5$  kcal/mol, whereas that of R190K/T192P was  $2.4$  kcal/mol and that of T192P was  $3.5$  kcal/mol. Again, the T192P mutation alone seems to affect function most profoundly, but is rescued to a degree by the R190K mutation. The  $\Delta\Delta G$  for T192P is approximately equivalent to the loss of one hydrogen bond, which correlates well with the loss of the bond between the Thr192 side chain hydroxyl group and the purine N7. Perhaps the Lys190 side chain offers greater flexibility to the backbone of the loop, allowing better accommodation of the Thr192Pro substitution yet still not allowing the corepressor binding pocket to form. It is not surprising that R190K/T192P protein does not bind hypoxanthine, as two residues that usually contact corepressor have been changed. This is particularly true if we assume the T192P mutation reorients Lys190 to make it less effective at contacting corepressor. Perhaps the binding pocket cannot form, or the Thr192Pro mutation disfavors the open (apo) form of the protein making the transition to the closed (holo) conformation.

Many of the ligand-binding residues in the ribose binding protein, RBP, are also conserved with those of PurR and, putatively, RbsR. The binding pockets of the two proteins are remarkably similar. Unlike RbsR, however, both PurR and RBP use arginine and threonine to bind their respective ligands, but in RBP the arginine comes from a different strand from that of the threonine. The residues that contact ligand are otherwise absolutely conserved, but for Tyr73-Phe74 that stack against the purine base in PurR, replaced by Phe15-Phe16 in RBP. Interestingly, in RbsR, these residues are present in the order Phe-Tyr. Perhaps this subtle rearrangement is necessary for conversion of the purine repressor to a ribose repressor. That RBP uses Arg and Thr to bind ribose in a manner similar to the way PurR uses Arg and Thr to bind purines may invalidate our theory that mutating these residues to align with the Lys and Pro of RbsR could alter corepressor specificity. Additionally, the *H. influenzae* purine repressor has a lysine at position 190 (Fleischmann et al., 1995).

Ligand specificity has been expanded or changed in a number of cases (reviewed in (Clackson, 1998)). In some examples, drastic changes have been made, such as the conversion of glutathione transferase A1-1 from a low activity and specificity enzyme that uses a nucleophilic substitution mechanism to a high activity enzyme that uses Michael addition with a specific set of nonenals (Nilsson et al., 2000). The transformation was made by substituting four residues and a portion of a helix from a different glutathione transferase protein. The specificity of the arabinose binding protein (ABP), which normally binds arabinose and galactose, was redesigned to allow binding of methyl- $\beta$ -D-galactopyranoside and isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) with only two amino acid residue substitutions (Declerck and Abelson, 1994). The mutant

protein no longer bound galactose and had ~3000-fold lower affinity for arabinose than wild-type ABP. In another study, the substrate specificity of trypsin was transformed to mimic that of chymotrypsin by replacing active site residues with those found in chymotrypsin, however two surface loops had to be substituted as well for complete transformation of specificity (Hedstrom et al., 1992). Interestingly, "rational protein design" is not always successful, as demonstrated with mutants of the well characterized EcoRV protein (Lanio et al., 2000). Several point mutations designed to allow the enzyme to discriminate between differentially flanked DNA recognition sites failed to do so, and structural analysis revealed that the mutated side chains did not take the predicted conformations.

In light of the extent of the changes required to make dramatic specificity changes in other proteins, it is possible that for PurR to bind ribose, the entire loop on which Arg190 and Thr192 reside, residues 185-193, must be replaced by the corresponding loop of RbsR. This loop contains two proline residues in wild-type PurR, and two in RbsR. Residues Gly186-Pro187-Lys188 are conserved between the two proteins. However, the second proline in PurR is at position 185, whilst the second proline in RbsR is at position 192. Positioning of the second proline at the beginning or end of the loop may be key to discrimination between purines and ribose. Substitution of the entire loop may shed light on the importance of the proline content and placement within this region.

## **Acknowledgments**

The R190K mutant was made by Fu Lu in the Zalkin laboratory. I carried out all other mutagenesis, binding and structural studies.

Table 5.1. Affinities of wild type and mutant purine repressors for *purF* operator alone or in the presence of hypoxanthine or ribose.

Protein	K <sub>d</sub>			
	+ DNA <sup>a, b</sup> + 2 mM hypoxanthine	+ DNA	+ DNA + 2 M Ribose	+ DNA <sup>c</sup>
Wild-type PurR	4.7 +/- 0.4 nM	NBD <sup>d</sup>	4.7 +/- 1.7 nM	1.5 +/- 0.3 nM
R190K	9.3 +/- 1.4 nM	NBD	8.9 +/- 4.6 nM	3.6 +/- 0.5 nM
R190K/T192P	NBD			93.0 +/- 11.6 nM
T192P	1.59 +/- 0.10 $\mu$ M			648 +/- 74 nM

<sup>a</sup> DNA used in all experiments was the 5'-fluoresceinated *purF* operator.

<sup>b</sup> Unless otherwise noted, all experiments were done at 35 °C in "corepressor-dependent" buffer (see Materials and Methods).

<sup>c</sup> These experiments were carried out at 25 °C in "corepressor-independent" buffer (see Materials and Methods).

<sup>d</sup> NBD stands for "no binding detected".

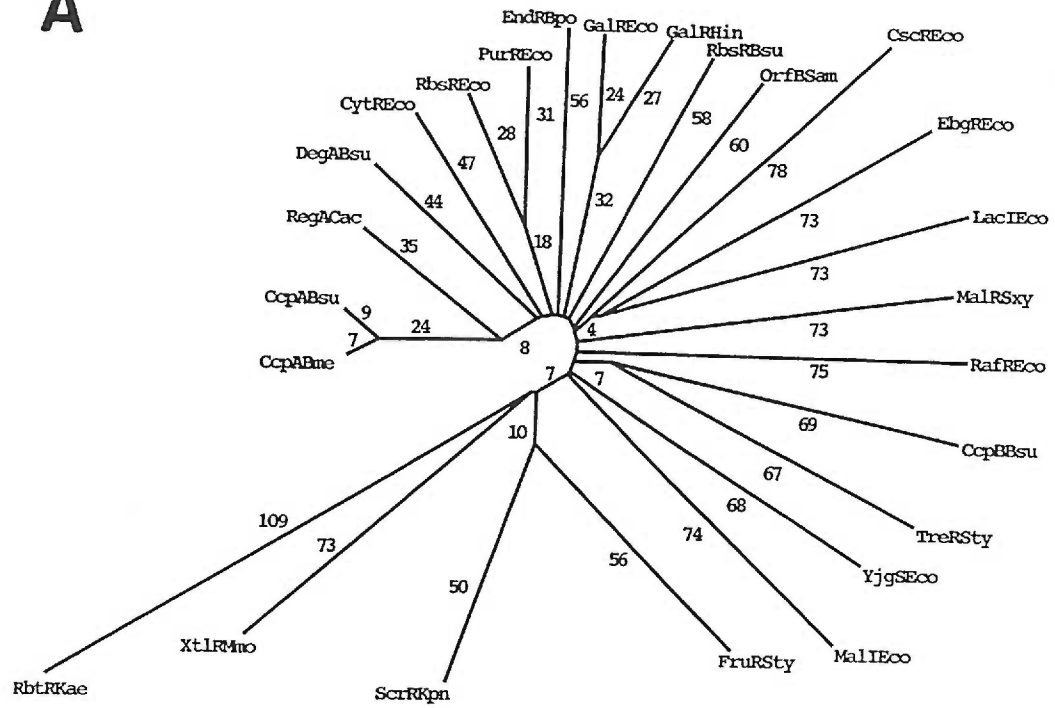


Table 5.2. Summary of selected crystallographic and refinement data for R190K-hypoxanthine-*purF* operator complex.

<i>Data Collection:</i>	
Resolution (Å)	30.0 - 2.7
Unique reflections (#)	16222
Completeness (%)	84.1
Rsym (%)	8.9
Temperature (°C)	22
Cell constants:	
a (Å)	175.45
b (Å)	93.94
c (Å)	81.57
$\alpha$ (°)	90.0
$\beta$ (°)	90.0
$\gamma$ (°)	90.0
<i>Refinement:</i>	
Resolution (Å)	10 - 2.7
R factor (%)	19.0
Total atoms (#)	3049
Water molecules (#)	34
rms deviations:	
bond length (Å)	0.014
bond angle (°)	1.54
B correlation (Å <sup>2</sup> )	3.77

Figure 5.1A. The LacI/GalR superfamily phylogenetic tree, adapted from (Nguyen and Saier, 1995). B. Primary sequence alignment of PurR and RbsR, generated using ClustalW alignment software (Thompson et al., 1994). Residues involved in corepressor binding are marked with an \*.

A



## B

```

PurR    1 MATIKDVAKRANVSTTTVSHVINKTRFVAEETRNAVWAAIKELHYSPSAVARSLKVNHTK
RbsR    1 MATMKDVARLAGVSTSTVSHVINKDRFVSEAITAKVEAAIKELNYAPSALARSLKLNQTH

          **
PurR    61 SIGLLATSSEAAYFAEIIIEAVEKNCFQKGYTLIGNAWNNLEKQRAYLSMMAQKRVDGLL
RbsR    61 TIGMLITASTNPFYSELVRGVERSCEFERGYSLEVCNTEGDEQRMNRNLETLMQKRVDGLL

PurR    121 VMCSEYPEPLLAMLEEVRIIPMVMDWGEAKADETDAVIDNAFEGCYMAGRYLIERGHRE
RbsR    121 LLCTETHQPSREIMQRYPTVPTVMMDWAPFDGD-SDLIQDNSLLGGDLATQYLIDKGHTR

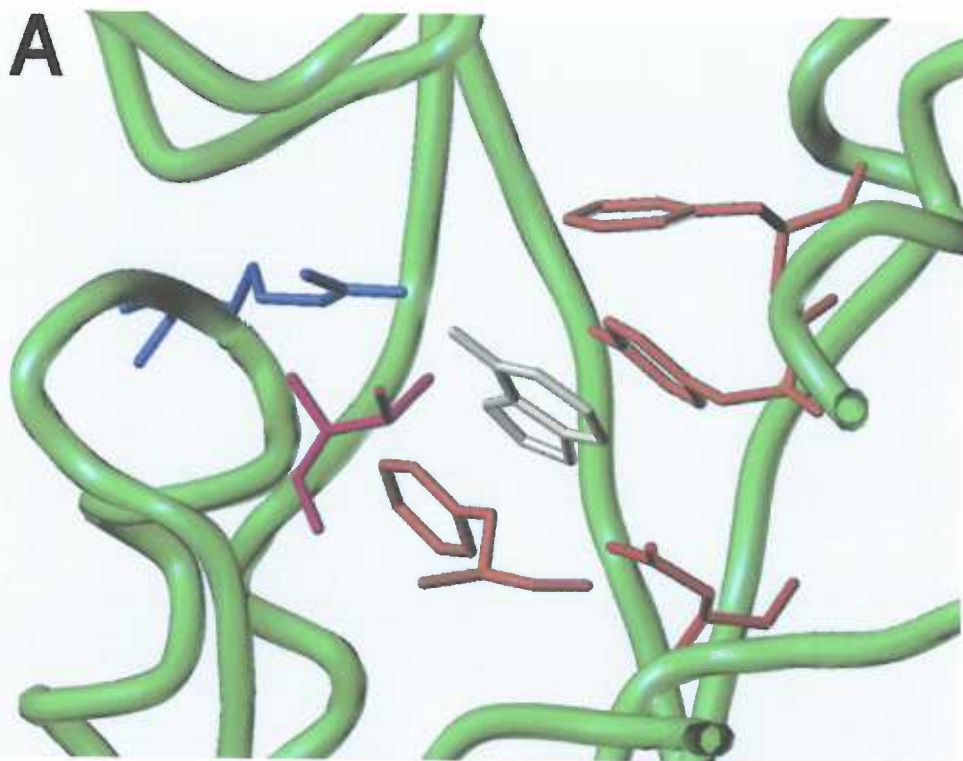
          * *
PurR    181 IGVIPGPLERNTGAGRLAGFMKAMEEAMIKVPESWIVOGDFEPESGYRAMQOILSQPHRP
RbsR    180 IACITGPLDKTPARLRLEGYRAAMKRAGLNIEDGYEVTGDFFENGGFDMARQLLSHPLRP

          *
PurR    241 TAVFCGGDIMAMGALCAADEMGLRVPQDVSLIGYDNVRNARYFTPALTTIHQPKDSLGET
RbsR    240 QAVFTGNDAMAVGVYQALYQAEIQVPQDIAVIGYDDIELASFMTPPLTTIHQPKDELGEL

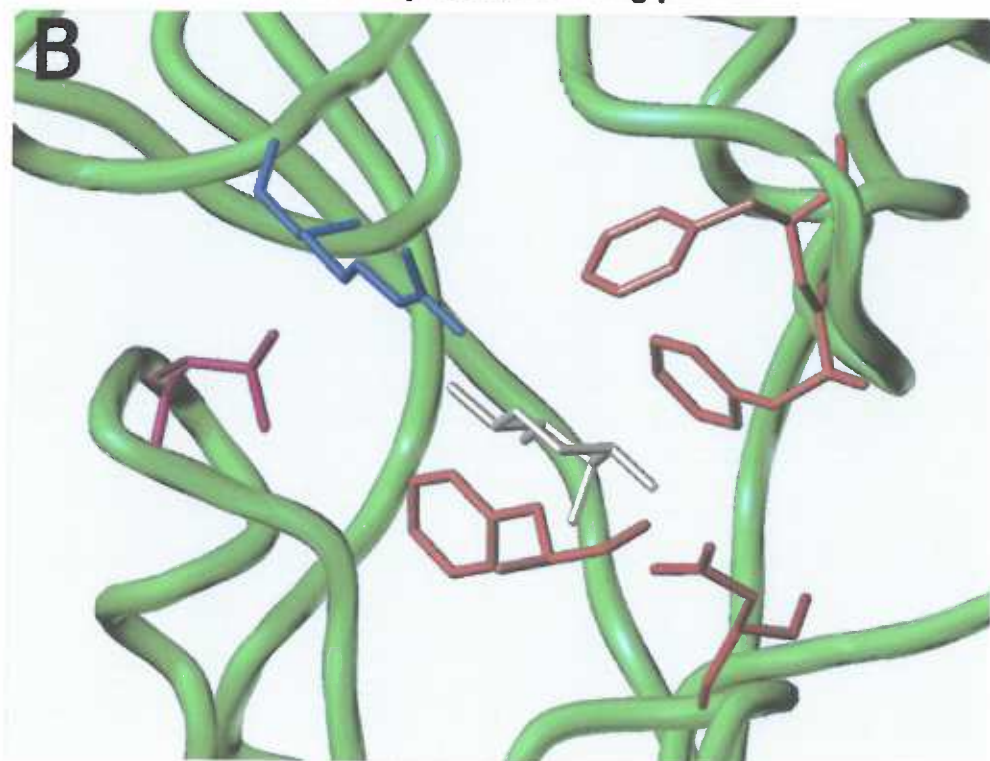
PurR    301 AFNMLLDRIVNKREEPOSIEVHERLIERRSVADGPFRDYRR 341
RbsR    300 AIDVLIHRITQPTLQQQRLQLTEILMERGSA----- 330

```

Figure 5.2. Close-up views of the ligand binding sites of A. PurR and B. RBP. The protein backbones are shown as green worms. In the PurR corepressor binding pocket, Arg190 is shown in blue, and Thr192 is shown in pink. Hypoxanthine is in white. Tyr73, Phe74, Phe221, and Asp275 are in orange. Residues in RBP are colored analogously, and ribose is in white.

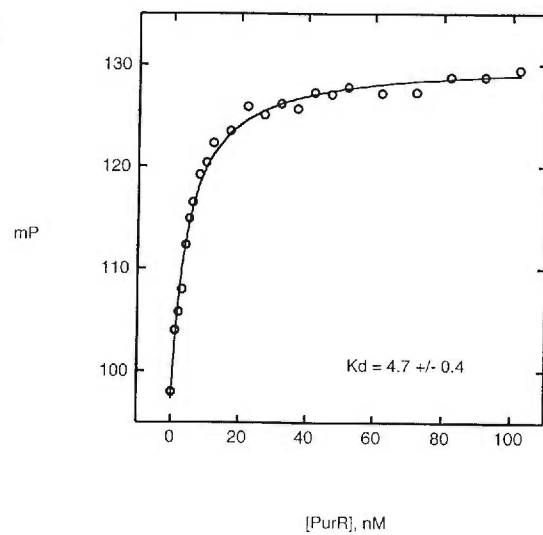
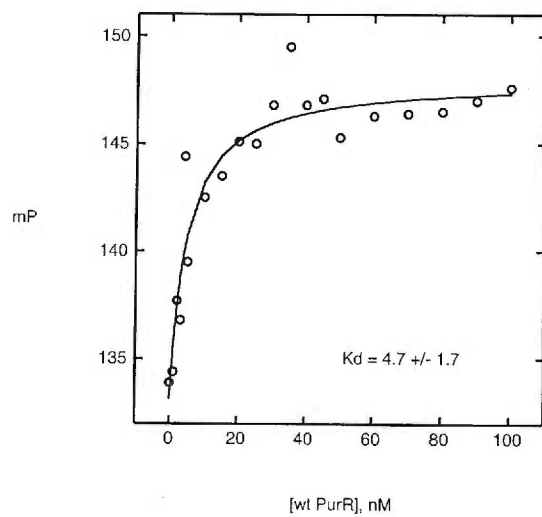
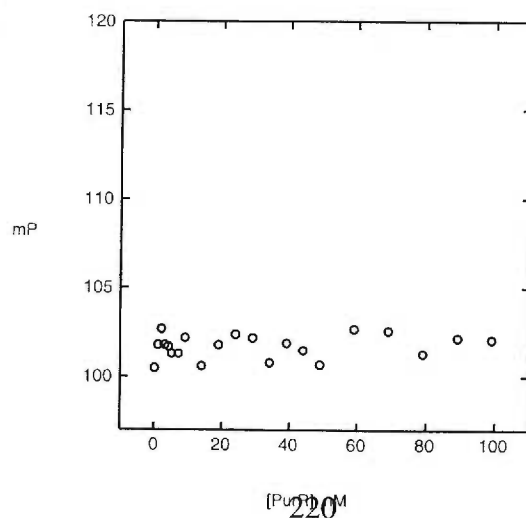


**PurR corepressor binding pocket**



**RBP ribose binding pocket**

Figure 5.3. Equilibrium binding isotherms for titrations of wild-type and mutant proteins into fluoresceinated *purF* operator and hypoxanthine, ribose, or no corepressor. *purF* operator is included in all titrations. A. wild-type PurR and hypoxanthine; B. wild-type PurR and 2 M Ribose; C. wild-type PurR and no corepressor; D. R190K and hypoxanthine; E. R190K and 2 M Ribose; F. R190K/T192P and hypoxanthine; G. T192P and hypoxanthine.

**A****B****C**



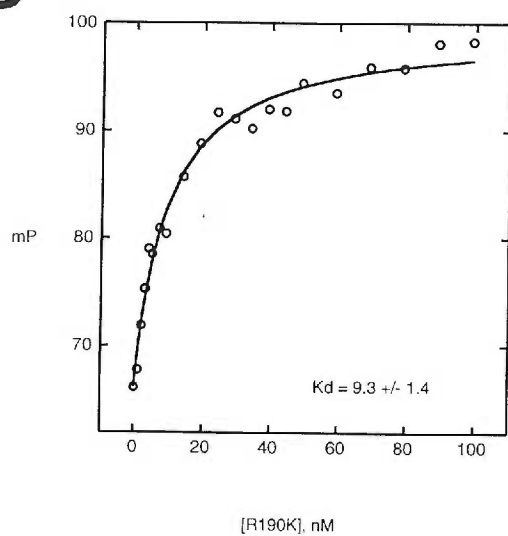
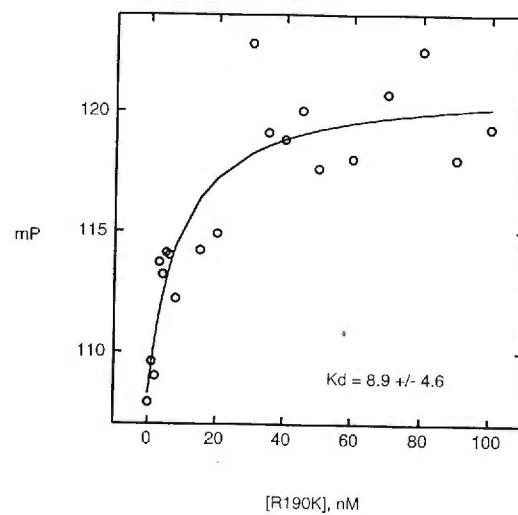
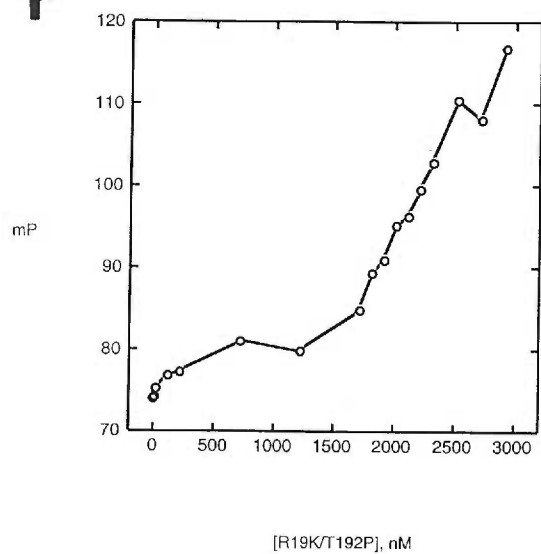
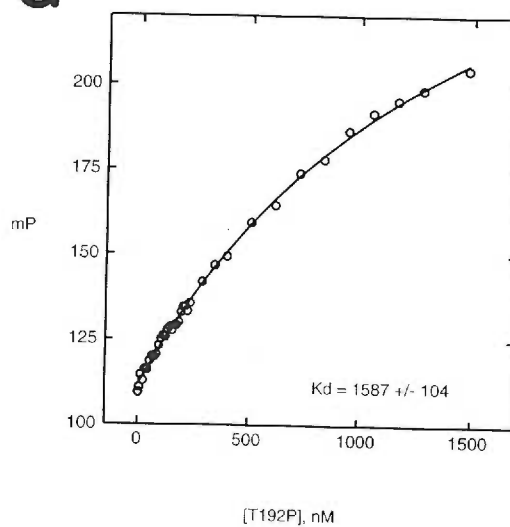
**D****E****F****G**

Figure 5.4. Equilibrium binding isotherms for titrations of hypoxanthine into fluoresceinated *purF* operator and A. wild-type PurR, B. R190K, C. R190K/T192P, and D. T192P.

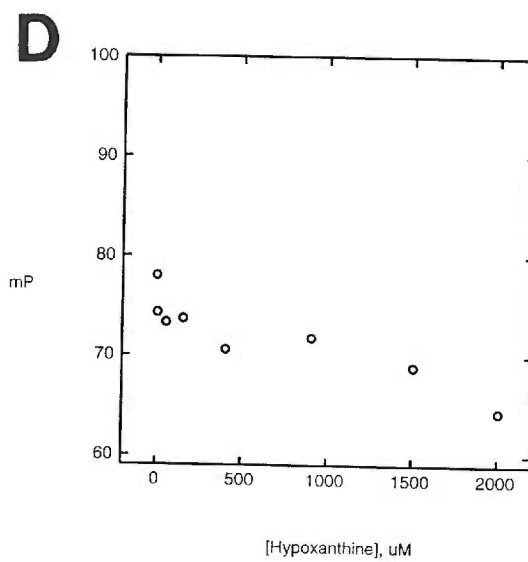
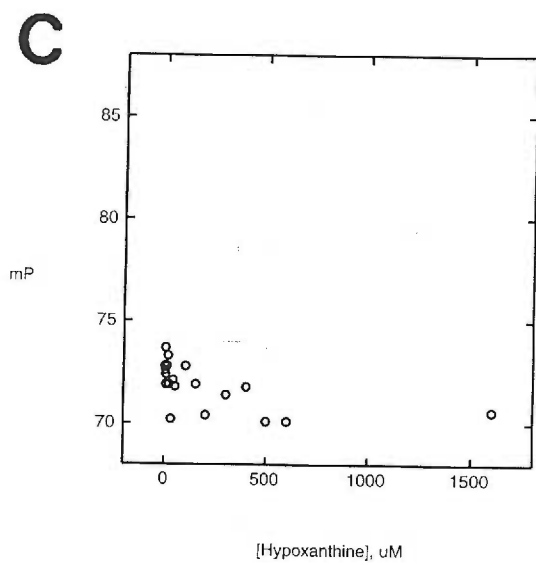
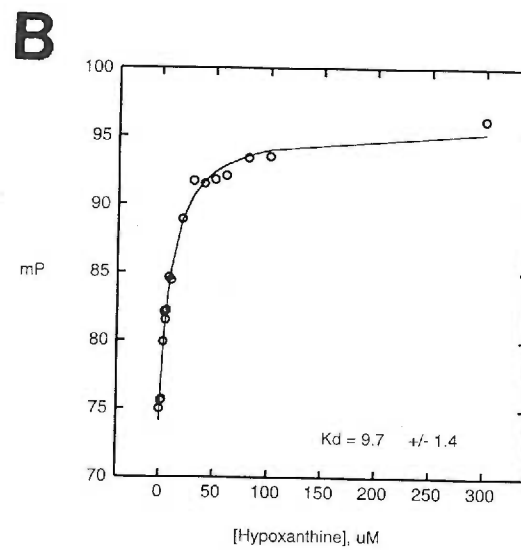
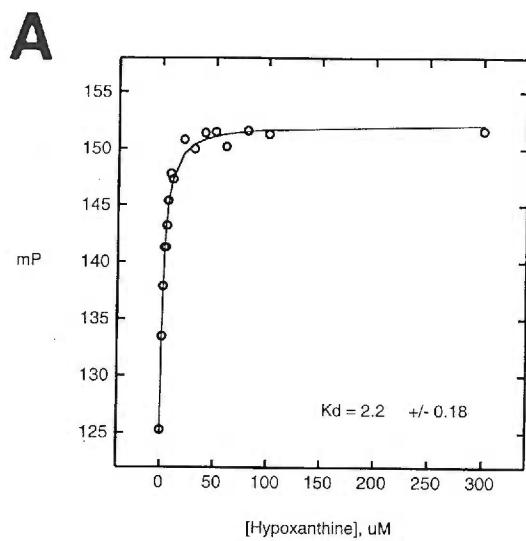


Figure 5.5. Equilibrium binding isotherms in corepressor-independent Buffer System I. A. Wild-type, B. R190K, C. R190K/T192P, and D. T192P proteins were titrated into fluoresceinated *purF* operator in a buffer that relieves the requirement for corepressor for efficient cognate DNA binding.

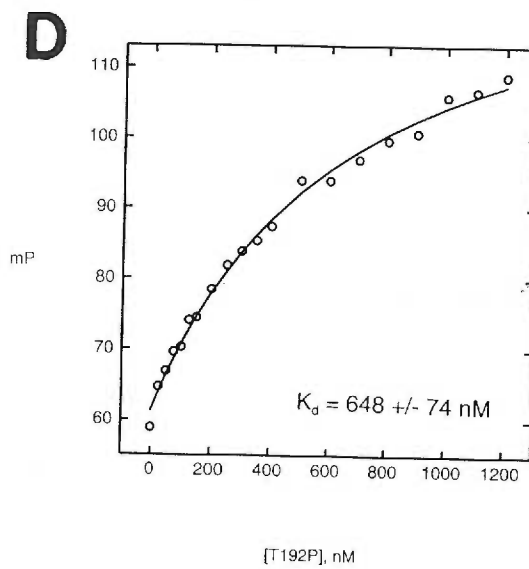
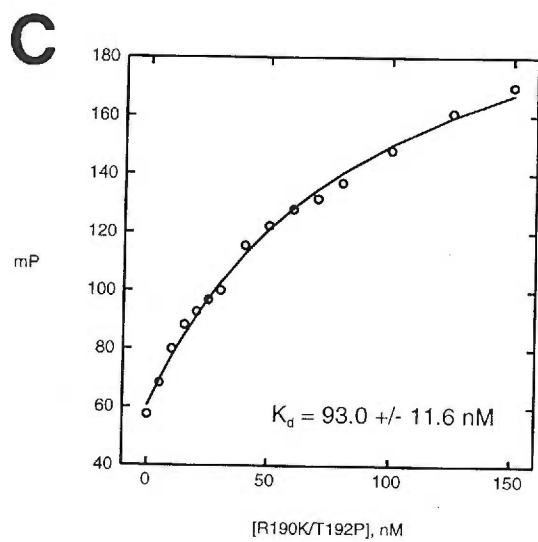
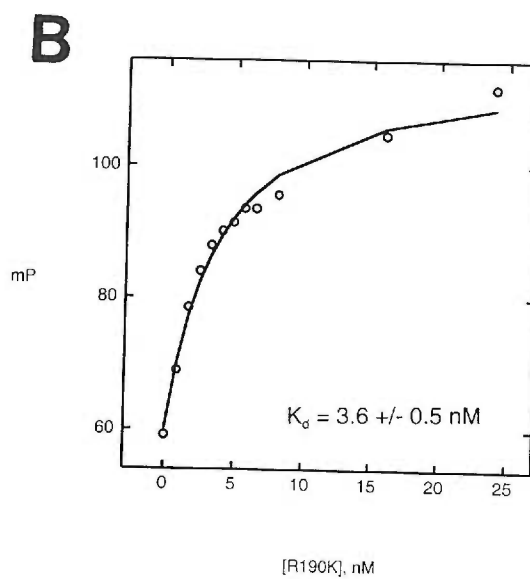
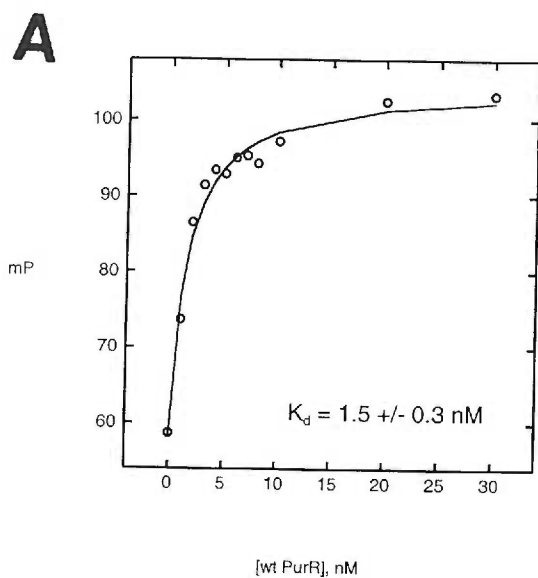


Figure 5.6A. Percent DNA binding plotted against ribose concentration for wild-type (▲) and R190K (○) proteins. B. Change in polarization (mP) upon addition of protein plotted against concentration of ribose for wild type (Δ), R190K/T192P (■), and T192P (□) proteins.

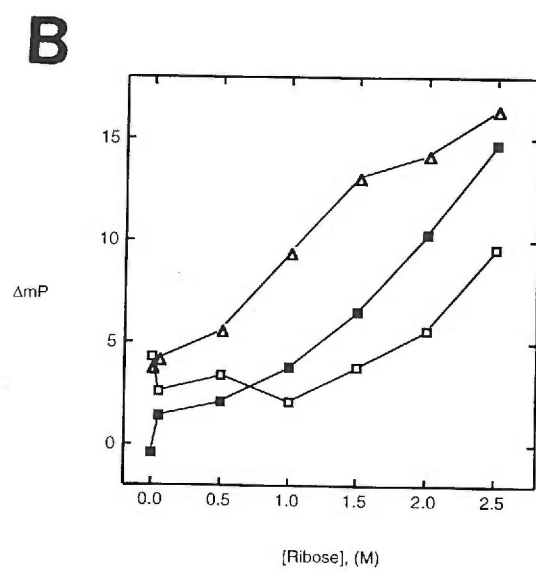
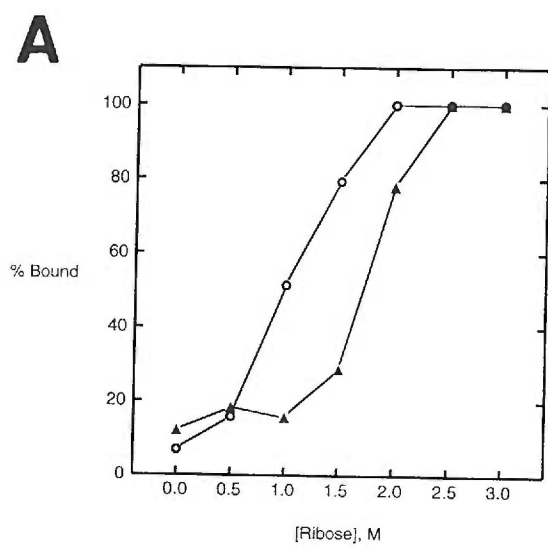


Figure 5.7. View of the corepressor binding pocket of the R190K mutant after the superimposition of their C $\alpha$  carbon atoms onto those of the wild type PurR protein bound to hypoxanthine. The R190K-hypoxanthine-*purF* operator is colored by atom type and the wild-type PurR-hypoxanthine-*purF* operator complex is in green. With the exception of those made by R190, all other wild type protein-purine contacts are found.



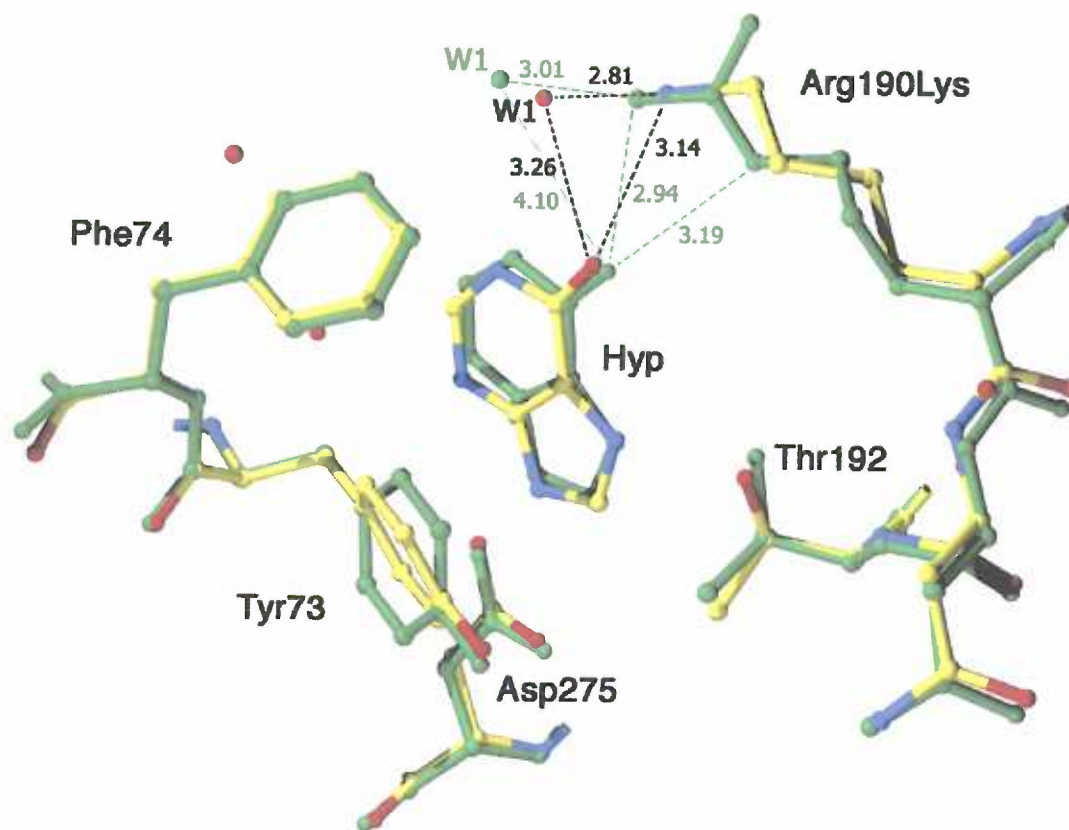
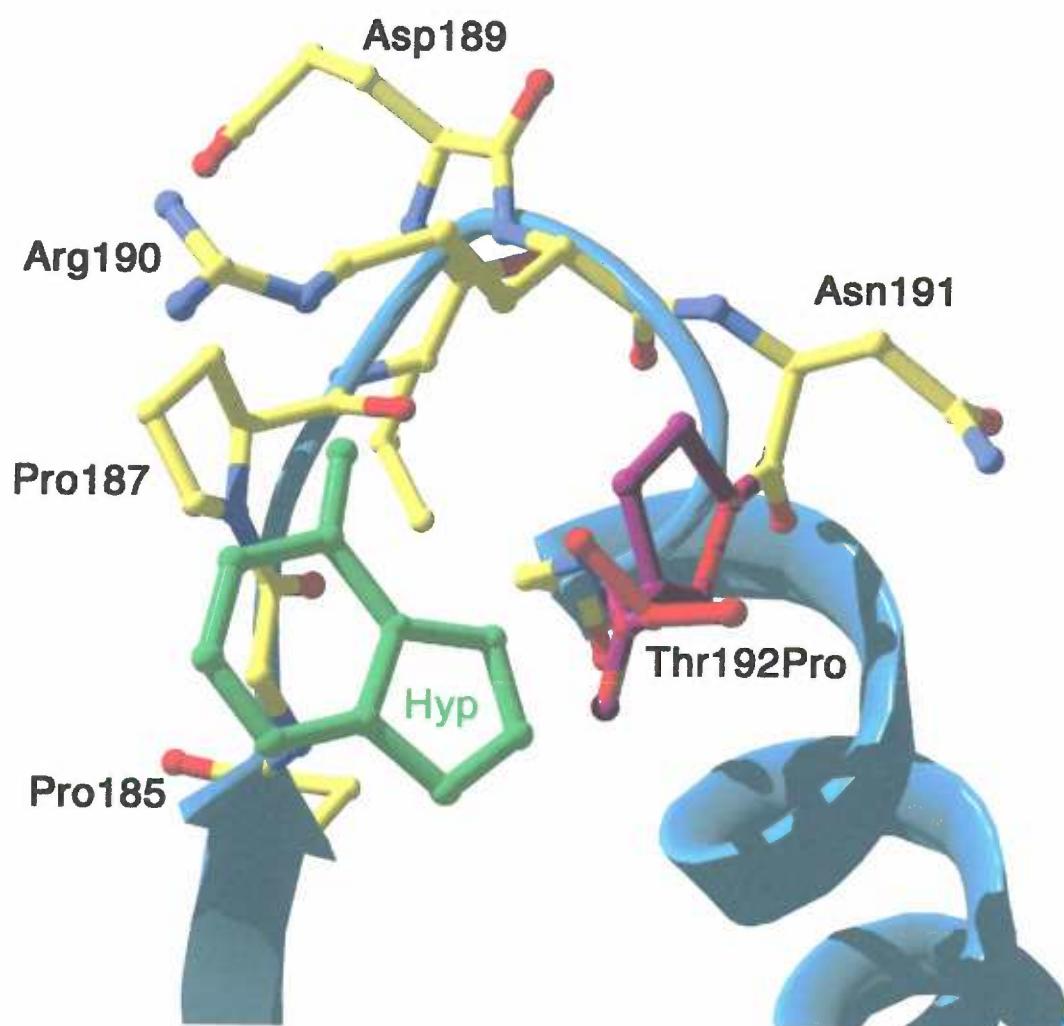


Figure 5.8. Model of the Thr192Pro mutation, based on the wild-type PurR-hypoxanthine-*purF* operator structure. The wild-type Thr192 is shown in red, and the Pro192 is in magenta. Hypoxanthine is in green.



**Chapter 6. DNA and metal specificity of MerR family member PMTR, *Proteus mirabilis* transcription regulator**

Joy L. Huffman<sup>1</sup>, Svetlana Lutsenko<sup>1</sup>, James R. Lundblad<sup>1,2</sup>, and Richard G. Brennan<sup>1</sup>

<sup>1</sup>*Department of Biochemistry and Molecular Biology and* <sup>2</sup>*Department of Medicine, Division of Molecular Medicine, Oregon Health Sciences University*

## Abstract

*Proteus mirabilis* transcription regulator, PMTR, is a member of the MerR family of prokaryotic transcriptional regulators that was found to confer resistance to normally toxic levels of zinc when transfected into *E. coli* cells. Here we report the determination of a cognate DNA binding site from *P. mirabilis* and show specific and high affinity *in vitro* binding of PMTR to this site. We have characterized the metal ion affinity of PMTR and show that it binds  $\text{Zn} > \text{Cd} \approx \text{Cu} \gg \text{Co}$ . DNA footprinting studies reveal that PMTR distorts its DNA binding site in a manner similar to that observed with other MerR family members, thus indicating a similar mechanism of transcription regulation.

## Introduction

Heavy metals are essential for cellular metabolism. Zinc, copper, cobalt, and nickel in particular are vital trace elements required for the functional activity and/or structural stability of many proteins (reviewed in (Coleman, 1992)). Cells must tightly regulate these metals within a narrow range since these metals are toxic at high concentrations. One basic mechanism of heavy metal resistance in prokaryotes is active efflux, resulting in reduced accumulation. Another mechanism is detoxification of the toxic ion. In prokaryotes, the activation of the efflux and detoxification pathway is regulated at the level of transcription. One class of regulators is the MerR family (Summers, 1992).

Members of the MerR family bind DNA as homodimers and activate or repress the transcription of genes involved in efflux or detoxification in response to the specific ligands to be transported (reviewed in (Summers, 1992)). All characterized members of this family bind promoter elements that have suboptimal spacing (18-20 base pairs, c.f. 17 preferred by RNA polymerase (Jacquet and Reiss, 1990) between the -35 and -10 regions, as was first demonstrated for MerR (Ansari et al., 1995; Frantz and O'Halloran, 1990). These transcription regulators bind cognate DNA in the presence or absence of ligand, typically repressing in the absence of ligand and activating when ligand has bound.

The MerR family is characterized by a conserved N-terminal DNA binding domain, followed by a long helical linker and a C-terminal effector binding domain. MerR family members bind a variety of specific ligands, reflected by the lack of homology found in their C-terminal effector binding domains. MerR binds Hg(II), and

several other heavy metal-dependent homologues have been reported that respond to increased levels of zinc, copper, silver, lead, and cobalt (Brocklehurst et al., 1999; Outten et al., 2000; Rutherford et al., 1999). SoxR is a MerR family member that responds to changes in intracellular oxidation and is activated by superoxides (Hidalgo et al., 1997). Other family members include BmrR, which binds multiple lipophilic cationic drugs, TipA, which binds the antibiotic thiostrepton, and Mta and BltR, whose endogenous ligands remain unidentified but appear to be involved in *B. subtilis* multidrug resistance (Ahmed et al., 1995; Baranova et al., 1999).

The *Proteus mirabilis* protein PMTR is a recently discovered member of this family, based on sequence alignments (Noll et al., 1998). The N-terminal domain of PMTR aligns well with the N-terminal DNA binding domain of BmrR and with those of other MerR family members. The PMTR C-terminal domain is much shorter than the C-terminal drug binding domain of BmrR, instead sharing its length and sequence conservation with the other metal binding MerR family members. PMTR transfected into *E. coli* cells was found to confer resistance to normally toxic levels of zinc. The exact mechanism of resistance is unclear, but correlates with the increased accumulation of a 12 kDa proteolytic fragment of the *E. coli* ZRAP protein in the periplasm. No function had previously been assigned to ZRAP, but it might have a role in the regulation of specificity of the heavy metal transporter, ZntA.

Since the discovery of PMTR, an endogenous zinc-responsive *E. coli* protein, ZntR, has been identified (Brocklehurst et al., 1999). The *E. coli* ZntR is a MerR family member that is activated by Cd(II), Pb(II), and Zn(II) cations (Binet and Poole, 2000). Another chromosomal *E. coli* MerR-like protein has been characterized recently, the

copper and silver-responsive CueR transcription regulator (Outten et al., 2000). Both ZntR and CueR regulate systems that confer resistance to the metals they bind. Their functional relationship to PMTR is unclear.

Our understanding of the structural mechanism by which a MerR family member regulates transcription was aided recently with the crystal structure of the *B. subtilis* multidrug binding protein, BmrR, bound as a homodimer to cognate DNA and a small molecule ligand, tetraphenylphosphonium (TPP) (Heldwein and Brennan, 2001). The mechanism of DNA distortion implied by footprinting experiments (Ahmed et al., 1994) and finally revealed in this structure is very unusual. The protein kinks the DNA by  $\sim 50^\circ$ , and the central region of the operator is unwound. Most surprisingly, the central AT base pair (bp) breaks its Watson-Crick hydrogen bonds, and the bases slide away from each other in the 3' direction. One of the overall effects of the untwisting and base sliding is the shortening of the operator by about 5 Å. This brings the suboptimally spaced -35 and -10 elements into phase on the DNA, thus allowing efficient recognition and utilization of both regions by RNA polymerase.

To analyze the DNA and metal binding properties of PMTR and its relationship to other metal binding MerR family members, we purified the protein and characterized possible DNA binding sites and their affinities. Electrophoretic mobility shift assays (EMSA) allowed the approximate identification of a DNA binding site, and footprinting experiments led us to precisely discern a specific PMTR binding site, which lies upstream of a putative *P. mirabilis* copper ATPase. Fluorescence polarization equilibrium binding assays revealed that PMTR specifically binds this DNA site with high affinity in the presence and absence of zinc. However, subtle differences in the footprint, analogous to



those observed with other MerR family members, are observed. We discuss our results in light of the structure of the BmrR-DNA-drug complex in its activated conformation (Heldwein and Brennan, 2001) and present a model of possible PMTR-DNA interactions.

## Materials and Methods

### *PMTR cloning, protein expression, and purification:*

The gene encoding PMTR (GenBank accession number # AF017747) was PCR-amplified from a plasmid containing a fragment of the *Proteus mirabilis* genome, designated PROT1 (Noll et al., 1998), using the following primers: forward 5' - tctgcacatatgaatattgggtcaagctgc - 3', reverse 5' - ggctctaagctttaatgttctgtccctg - 3'. These primers add a HindIII restriction site to the 5' end of the PCR product and an NdeI site to the 3' end, allowing in-frame subcloning into the pET15b expression vector (Novagen) such that the PMTR protein produced contains six N-terminal histidine residues for ease of purification. The resulting plasmid was transformed into *E. coli* BL21λDE3 cells. To produce protein, cultures of these cells were grown in LB and 100 µg/ml ampicillin at 37 °C until mid-log phase was reached, at which time the cultures were transferred to 25 °C, and PMTR protein production was induced with 0.5 mM IPTG. Cells were harvested after four hours. The temperature change at the time of induction resulted in a larger fraction of PMTR protein in the soluble fraction of cell lysates. Cleared cell lysates were passed over Ni-NTA resin (Qiagen), which was then washed extensively with Buffer I (30 mM imidazole, 200 mM NaCl, 50 mM Tris, pH 7.6, 5% glycerol). PMTR protein was eluted from the Ni-NTA resin using a 30 to 300 mM imidazole gradient. This protein was dialyzed against its storage buffer, 200 mM NaCl, 5 mM DTT, 50 mM Tris, pH 7.6.

### *Radiolabelling of DNA for footprinting and gel shift experiments:*

Two different 200 base pair fragments of DNA were PCR-amplified from the PROT1 plasmid (Noll et al., 1998). This plasmid harbors a 13.5 kb genomic fragment of *P. mirabilis* DNA, containing at least four genes (Figure 6.1A). These are: *hmpA* and *hmpB* (hemolysin genes required for *P. mirabilis* virulence), *atpase* (a gene encoding a putative copper-transporting ATPase), and the *pmtr* gene, encoding a putative transcription factor of the MerR family (Figure 6.1). The first fragment, called *atpaseP*, contained the 151 bp intergenic region between the *hmpB* and *atpase* genes and 25 or 24 bp of flanking sequence from the plasmid on the 5' and 3' sides, respectively (Figure 6.1B). This region was amplified using the following primers: template strand 5' - cggaattcatggcatctgctcccgataaac - 3', non-template strand 5' - gcggatccggataatgtttaggtgtattc - 3'. The second fragment, called *pmtrP*, contained the 111 bp intergenic region between the *atpase* and *pmtr* genes and 46 or 43 bp of flanking sequence from the plasmid on the 5' and 3' sides, respectively (Figure 6.1B). The primers used to amplify this region had the sequences: template strand 5' - gcaagcctaataatgcactacgactaaaacg - 3', non-template strand 5' - cggaattctacttgaaaataccggatttttg - 3'. Primers were 5' labelled with T4 polynucleotide kinase (NEB) and [ $\gamma^{32}\text{P}$ ]-ATP (NEN, >6000 Ci/nmole). One  $^{32}\text{P}$ -labelled primer and one unlabelled primer were used in PCR reactions to amplify fragments in which either the template or non-template strand was 5'  $^{32}\text{P}$ -labelled. Fragments were purified by agarose gel electrophoresis, phenol and chloroform extraction, and ethanol precipitation and resuspended in TE buffer (1 mM EDTA, 10 mM Tris, pH 8.0). The template strand-labelled DNA will be denoted with an F, for forward, as in *atpaseP*-F, and the non-template strand-labelled DNA will be denoted R, for reverse, as in *atpaseP*-R.

*Electrophoretic mobility shift assays (EMSA).*

EMSA were performed with each radiolabelled PCR product: *atpaseP-F*, *atpaseP-R*, *pmtr-F*, and *pmtr-R*. In 10 µl reactions, 1 µl radioactive probe DNA (~0.5 pmoles) was mixed with 2 µl 5x binding buffer (250 mM NaCl, 50 mM Tris, 7.6, 25% glycerol), 1 µg poly d(IC), 0.5 µg BSA, and titrations of 0, 0.01, 0.05, 0.1, 0.2, 0.5, and 1.0 µM PMTR protein. Reactions were incubated for 30 minutes at 25 °C then loaded onto 4% acrylamide/TGE (25 mM Tris, 192 mM glycine, 1 mM EDTA, pH 8.5) gels and electrophoresed for 30 minutes at 100 V, using TGE as the running buffer. Gels were dried and put against a K-screen (Kodak) for 20 minutes. This screen was then scanned at 100 µm resolution using a Molecular Imager Fx phosphorimager (Biorad).

*DMS methylation protection experiments.*

DNA sequencing reactions were performed using the method described by Maxam and Gilbert (Maxam and Gilbert, 1977), with minor modifications. Dimethylsulfate (DMS) methylation protection reactions were similar to the Maxam-Gilbert guanine sequencing reaction, with some modifications. Briefly, samples were incubated prior to reaction as described for gel shift assays, but with 5 µl probe DNA ( $\leq 0.5$  µM) and 15 µl of either buffer or 10 µM PMTR protein in a 20 µl volume, to which 30 µl 1:200 DMS in cacodylate buffer was added. The reaction was terminated with 200 µl DMS stop buffer (300 mM Na acetate, 0.8%  $\beta$ -mercaptoethanol, 0.025 mg/ml yeast tRNA). Maxam-Gilbert sequencing reaction protocols were then followed (Maxam and Gilbert, 1977), in which the DNA was precipitated with ethanol, washed, and then

cleaved with piperidine. The final DNA pellet was resuspended in formamide loading buffer (80% formamide, 1% NaOH, 0.1 mM EDTA, 0.1% dye) and loaded onto a 4% acrylamide/TBE (Tris-borate-EDTA) sequencing gel. Sequencing gels were run at 50 W for 3 to 4 hours. Gels were dried and analyzed by phosphorimager as described for EMSA.

#### *DNaseI protection footprinting.*

PMTR-DNA binding reactions were carried out under the conditions of gel shift experiments, but adjusted to a volume of 50  $\mu$ l. Two separate reactions were performed, with and without 250  $\mu$ M ZnCl<sub>2</sub>. After a 30' incubation of DNA and protein-DNA solutions, 6  $\mu$ l of 10x salt solution (50 mM MgCl<sub>2</sub>, 10 mM CaCl<sub>2</sub>, 20 mM dithiothreitol) were added, followed by 1.25  $\mu$ l of 1:100 DNaseI stock (0.0125 units). After 2 minutes, the reaction was stopped with 350  $\mu$ l DNaseI stop solution (7% saturated ammonium acetate, 7  $\mu$ g/ml tRNA, 93% ethanol). The DNA was extracted with phenol and chloroform, precipitated with ethanol, and washed with cold 70% ethanol. The resulting DNA pellet was resuspended in formamide loading buffer, and loaded onto a sequencing gel, which was run and analyzed as described in the previous section.

#### *Hydroxyl radical cleavage reactions.*

PMTR-DNA binding reactions were carried out exactly as with DNaseI treatments, but PMTR was first dialyzed overnight into storage buffer without glycerol, and the binding buffer contained no glycerol for these experiments. Glycerol is a potent free radical scavenger and thus competitively inhibits hydroxyl radical cleavage of DNA.

DNA cleavage reactions were carried out according to standard methods (Dixon et al., 1991), with minor modifications. To the side of the Eppendorf tubes containing the DNA-buffer or DNA-PMTR mixtures, equal amounts of 2.5 mM Fe[EDTA]<sup>2-</sup>, 10 mM sodium ascorbate, and 0.3% H<sub>2</sub>O<sub>2</sub> were added. The tubes were centrifuged to mix the components. To stop the reaction, 0.1 M thiourea and 10 mg/ml tRNA were added to the sides of the tubes, which were centrifuged to mix. Water was added, and the DNA was extracted with phenol and chloroform then precipitated with ethanol. The DNA pellet was washed with 70% and then 100% ethanol, dried, and resuspended in formamide loading buffer. This was then loaded onto a 4% acrylamide/TBE sequencing gel. This gel was electrophoresed and analyzed as described for DMS methylation experiments.

#### *Fluorescence polarization.*

5'-Fluoresceinated duplex (or stem-loop) oligonucleotides were used in all fluorescence polarization experiments. Sequences for the top (or entire, for stem-loop) strands are as follows:

*pm-atp* 5' - ataatccttgaccttcccctaattggttaaagcttaagctttgt - 3',

*ec-znt* 5' - taaaacttgactctggagtcgactccagagtgtatccttcgg - 3',

stem loop *ec-purF* 5' - aaagaaaacgtttgcgtacccctacgcaaacgttttctt - 3'.

*Pm-atp* includes the region of the *atpaseP* promoter identified in footprinting experiments to be bound by PMTR. *Ec-znt* is the analogous operator region specifically bound by the *E. coli* ZntR protein, another MerR family member (Outten et al., 1999). *Ec-purF* is an unrelated operator bound by the *E. coli* purine repressor. Oligonucleotides were purchased from Oligos, Etc. Template and non-template strands were annealed to

form duplex DNA by mixing equimolar amounts of each strand in 10 mM sodium cacodylate, pH 6.5, heating to 80 °C for 10 minutes, and allowing the solution to cool slowly to room temperature. Stem-loop structured DNA was self-annealed in the same DNA storage buffer by heating to 80 °C for 10 minutes then flash-cooling the solution by immediately placing it on ice.

All fluorescence polarization experiments were carried out in 50 mM NaCl, 20 mM Tris, pH 7.6, 1 µg/ml poly d(IC), and 0 or 250 µM ZnCl<sub>2</sub>. PMTR protein was titrated into a 1 ml solution of buffer and 2 nM fluoresceinated DNA.

The data of each binding isotherm were analyzed by nonlinear regression using SigmaPlot software (Jandel Corporation). Because the dissociation constants were all above 20 nM and the experimental DNA concentration was ten-fold less, it was assumed that the concentration of protein bound to DNA was negligible in comparison to the total protein concentration. Therefore, the following equation could be applied:

$$A = \left\{ \frac{(A_{\text{bound}} - A_{\text{free}})[P]}{K_d + [P]} \right\} + A_{\text{free}}$$

where A is the polarization measured at a given total concentration of protein [P], A<sub>free</sub> is the initial polarization of the free DNA, and A<sub>bound</sub> is the polarization of maximally bound DNA (Glasfeld et al., 1996; Lundblad et al., 1996). Least squares regression analysis was used to determine A<sub>free</sub>, A<sub>bound</sub>, and K<sub>d</sub>.

#### *Metal specificity determination.*

To test the specificity of metal binding by PMTR, 2 µM purified PMTR protein in 30 µl was incubated with 100 µM Zn, Cd, Co, or Cu chloride salts, or with PMTR

storage buffer. After 15 minutes, 1  $\mu$ l of 1 mM fluorescent coumarine maleimide CPM (Molecular Probes), a cysteine labelling reagent, was added to the incubation mixture. Following a 2 minute incubation at ~22 °C in the dark, the reaction was stopped by addition of Laemmli sample buffer (2.7 M urea, 3.3% SDS, 0.33 M Tris-HCl, pH 6.8) containing 1%  $\beta$ -mercaptoethanol. The protein was loaded onto a 12% acrylamide Laemmli gel and electrophoresed. Labelled protein was visualized using a GelDoc instrument (Biorad) and quantified using Molecular Analyst software (Biorad) .



## Results

*Identification of a PMTR binding site within the atpase promoter but not the pmtr promoter.*

We used EMSA to test the binding of PMTR to two 200 base pair (bp) DNA fragments containing the putative promoter regions of the genes encoding a possible Cu-ATPase (*atpase*) and PMTR itself (*pmtr*), illustrated in Figure 6.1. *AtpaseP*-F and *atpaseP*-R both correspond to *atpaseP* DNA <sup>32</sup>P-labelled at the 5' end of either the template (F) or non-template (R) strand. *PmtrP*-F and *pmtrP*-R are also equivalent. Identical binding is expected with F- and R-labelled strands. A high affinity shift is observed with *atpaseP* DNA, corresponding to the putative promoter of the ATPase (Figure 6.2A). At protein concentrations as low as 10 to 50 nM, a single shifted band is observed. At successively higher concentrations, a second and a third band are observed, similar to the two shifts seen at high concentrations with *pmtrP* (Figure 6.2B). *PmtrP* DNA was shifted only at higher PMTR concentrations. However, unlike the *atpaseP*-containing fragment, the shifted bands do not coalesce into a single band at 3 μM PMTR. Additionally, gels were run in Tris-glycine-EDTA buffer, which may alter the specificity of the protein for DNA. Under our experimental conditions, the results do not indicate cooperative binding, rather that tetramerization or higher oligomerization occurs at protein concentrations above 10 nM.

To identify the specific site(s) bound by PMTR and to visualize any characteristic MerR-family DNA distortion caused by PMTR binding, we initiated a series of DNA cleavage protection, or "footprinting" experiments in the presence or absence of PMTR

protein and in its Zn bound and Zn free conformations, in which the DNA was subjected to methylation protection or DNaseI and hydroxyl radical cleavage.

#### *Methylation protection.*

A 200 nucleotide fragment containing the *atpaseP* DNA was labelled on either the template (*atpaseP-F*) or the non-template (*atpaseP-R*) strand. Upon inclusion of PMTR protein, a clear change in DMS reactivity was seen (Figure 6.3, lanes 5, 6, 19, 20). The results are summarized in Figure 6.5, which shows that in the presence of PMTR and without zinc, two guanine bases are hypersensitive (red asterisks), one on each strand (template -62 and non-template -53), and six guanines are protected from methylation by the PMTR (template -48, -47, -42, -36, non-template -60 and -59, green asterisks), as inferred from the difference in reactivities of the same guanines in the absence of protein. DMS cleavage reactions were not done in the presence of zinc.

#### *DNaseI interference footprinting.*

DNaseI interference experiments gave a pronounced footprint in the presence of PMTR (Figure 6.3, lanes 7-10, 21-24). This footprint extended from position -36 to -70, as noted in Figure 6.5 (blue lines). Subtle differences in protection were observed if zinc was included. DNA incubated with both PMTR and zinc shows a loss of protection of two bases on the non-template strand (-68 and -67), and two bases on the template strand (-67 and -66). This suggests that the protein-DNA interface is changed or that the DNA structure, and thus the availability of certain bases for DNaseI cleavage, has been altered by a change in protein conformation. PMTR provided no specific protection to *pmtrP-F*

or *pmtrP*-R DNA (data not shown). All further studies reported are with *atpaseP*-F or *atpase*-R DNA.

#### *Hydroxyl radical interference footprinting.*

The results of hydroxyl radical interference footprinting are, as anticipated, more difficult to visualize by sight, however they also yield a higher resolution footprint than the bulky DNaseI enzyme due to greater accessibility to the DNA. The data are shown in Figure 6.3, lanes 11-14 and 25-28. The linear traces of radioactivity down the gel in each of the lanes, given to easier interpretation, are plotted in Figure 6.4. The results are summarized in Figure 6.5, together with the results of the DMS and DNaseI experiments.

On the template strand, the PMTR protection footprint in the absence of zinc is clearly visible from -59 to -56, -51 to -47, and -41 to -39. On the non-template strand, we see a clear PMTR footprint from -68 to -67 and -64 to -41 in the absence of zinc. The protection is stronger at regions -64 to -61, -56 to -49, and -45 to -42. In the presence of zinc, subtle differences are observed. The template strand is protected at four additional bases, -54 to -52 and -42, by PMTR and zinc. The regions -41 to -39 and -49 to -51 have increased protection in the presence of zinc. The non-template strand has weaker protection of bases -68, -67, -60, -56, and -52 relative to protection observed in the absence of zinc and PMTR. The remaining interactions are similar to those observed without zinc.

#### *DNA Binding by fluorescence polarization.*

The results of our protein studies indicated that a 36 bp fragment, which encompassed promoter region -36 to -70, was specifically bound by PMTR. To determine whether this binding occurred at physiologically relevant concentrations and whether zinc altered the affinity for this site, we carried out a series of fluorescence polarization based binding assays, which allowed us to quantify the *in vitro* affinity of the *P. mirabilis* PMTR protein for this fragment and for nonspecific DNA. *Pm-atp* is a 42-mer oligonucleotide that includes the maximal region footprinted by DNaseI, with nucleotides of natural sequence flanking each side. The resulting binding isotherm is shown in Figure 6.6A, and the calculated equilibrium dissociation constant,  $K_d$ , is 20.8 +/- 3.9 nM. In the presence of 250  $\mu$ M  $ZnCl_2$ , we see virtually identical binding, with  $K_d$  equal to 26.9 +/- 3.9 nM (Figure 6.6B).

The specificity of PMTR for *pm-atp* was confirmed by comparing the binding isotherms of PMTR titrated into *pm-atp*, *ec-znt*, and *ec-purF*. *Ec-znt* is the operator recognized by *E. coli* ZntR, and *ec-purF* is the unrelated *purF* operator recognized by the *E. coli* purine repressor, which is not a member of the MerR family. The binding isotherm for *ec-znt* is shown in Figure 6.6C. The  $K_d$  was calculated to be 64.0 +/- 4.9 nM. Figure 6.6D shows the binding isotherm for PMTR binding to *ec-purF*. This binding isotherm is linear beyond 150 nM protein, and a  $K_d$  could not be estimated reasonably. However, the  $K_d$  value obtained by curve fitting was mM.

#### *PMTR metal specificity.*

The specificity of PMTR for various metals was tested using an *in vitro* labelling experiment. MerR uses three cysteines to bind its heavy metal ligand, Hg(II) (Helmann

et al., 1990). Two of these three cysteines are conserved in PMTR, Cys124 and Cys134 (Figure 6.8B), and PMTR contains one additional cysteine at position 52. We employed a fluorescent cysteine-labelling reagent, CPM, to test the availability of cysteines in the presence of various metals in order to determine the specificity for those metals. The results of the CPM labelling experiment are shown in Figure 6.7. In Figure 6.7A, the PMTR protein is clearly labelled differentially in the presence of different metals, while in Figure 6.7B the amount of protein loaded in each lane is observed to be nearly equal. The quantification of CPM labelling is presented in Figure 6.7C, showing that zinc gave the highest protection from the cysteine-labelling reagent, followed by copper and cadmium, with no protection observed in the presence of cobalt. These results verify those obtained previously with PMTR in *in vivo* experiments (Noll et al., 1998), and validate our use of zinc in the DNA binding experiments reported here. We infer that at least one of the three cysteines in PMTR is involved in binding metal, although it is possible that binding metal changes the availability of a cysteine without direct coordination.

## Discussion

*P. mirabilis* is the second most frequently isolated Enterobacteriaceae species in hospitals after *E. coli* (de Champs et al., 2000), and it is the cause of several types of urinary tract infection (Mobley et al., 1994). Mechanisms of heavy metal resistance have not yet been elucidated for *P. mirabilis* to the extent that they have in *E. coli*. One mechanism of zinc resistance has been proposed (Rensing et al., 1998), involving a Zn(II)-translocating P-type ATPase.

We have determined that PMTR binds the *P. mirabilis atpase* promoter at a specific site with high affinity (~21 nM under our *in vitro* conditions). We also show that upon addition of zinc, the protein-DNA interface is changed, as indicated by the altered DNA cleavage protection pattern obtained in the presence of zinc. The DNA footprinting results reported here correlate well with those obtained for other MerR family members (Ansari et al., 1995; Frantz and O'Halloran, 1990; Outten et al., 1999; Stoyanov et al., 2001; Summers, 1992). PMTR appears to share the DNA distortion mechanism common to MerR family members. MerR family members activate transcription in the presence of ligand, however not all family members repress in the absence of ligand. MerR represses transcription 2 to 10 fold in the absence of mercury, however, another family member, TipAL, activates transcription in both the absence and presence of the inducer, thiostrepton (Holmes et al., 1993). Thiostrepton addition induces expression over 200 fold.

We also observed specific binding with moderate affinity (~64 nM) to the *E. coli zntA* promoter, which has been shown to be bound *in vivo* and *in vitro* by another MerR family member, the *E. coli* ZntR protein. This binding is not wholly unexpected, as

previous experiments have shown that PMTR can modulate zinc resistance in *E. coli* (Noll et al., 1998). It is possible that PMTR increases zinc tolerance in *E. coli* through its interaction with the *zntA* promoter. However, we do not detect PMTR-induced production of the ZntA protein *in vivo* by immunoaffinity experiments (data not shown). It may be that another, similar *E. coli* operator is involved. A BLAST search for homologous operator sites to the *P. mirabilis atpaseP* site described in this study returned many regions upstream of genes with no assigned function, several of which were from *E. coli* (data not shown). Thus, the mechanism of zinc resistance in *E. coli* that is induced by PMTR remains to be elucidated. However, we found very high identity (75%) of our sequence with the operator recognized by *E. coli* CueR, a copper and silver binding MerR family member (Figure 6.8A, Table 6.1, (Outten et al., 2000)). The entire left pseudodyad is conserved between the promoters. This similarity is quite high considering that the *E. coli zntA* operator sequence shares only 15% identity with that of *atpaseP* (Table 6.1, (Outten et al., 1999)). *In vivo* repression assays have been initiated to clarify the activity of PMTR at various promoter regions.

Sequence alignment of PMTR, CueR, and ZntR reveals that all three proteins have high sequence identity (34-37%, Table 6.2). PMTR and CueR share conserved features that are distinct from other MerR family members (Figure 6.8B, Table 6.2). While the DNA binding domains (residues 1-69 of PMTR) of each of the MerR family proteins listed in Table 6.2 are well conserved (19-47% identity with an average of 35%), the long helical linker and effector binding domains are less well conserved (7-31% identity, with an average of 16%). Yet, within this region, PMTR and CueR are 39% identical, while PMTR and ZntR are 24% identical. Three cysteines at positions 84, 124,

and 134 (Figure 6.8B) are known to coordinate mercury in MerR (Helmann et al., 1990). Although these three cysteines are conserved in ZntR, only the last two are present in PMTR and CueR. Instead of a cysteine at position 76, both proteins have a serine in a short region of conserved sequence R(T/H)SADVK. The two cysteines at positions 112 and 120 are separated by only seven amino acids in PMTR and CueR, whilst the MerR spacer is eight amino acids, and the distance is nine residues in ZntR. These subtle differences may reorient the ligand binding pocket enough to allow binding of different metals. We have found that purified PMTR is able to bind zinc, cadmium, and copper *in vitro*, and, interestingly, neither ZntR nor CueR matches this metal-specificity profile

The structure of the BmrR-DNA-drug complex provides insight into the mechanism of activation by MerR family members, but it can also serve as a template on which to model protein-DNA interactions for other family members. PMTR has good sequence homology with BmrR in the N-terminal DNA binding domain (28% identity), but there is very little homology between the respective linker and effector binding domains (only 7% identity). The PMTR metal binding domain is also much smaller (~66 residues) than the BmrR drug binding domain (~200 residues). For this reason, in constructing a model of PMTR based upon the three-dimensional structure of BmrR, we have only used information from BmrR about the DNA binding domain. Shown in Figure 6.9 is a model of the dimeric PMTR-DNA interaction. Figures 6.10A and B are close-up views of the DNA binding motif likely adopted by PMTR. Comparing this constructed model to the footprinting results summary illustration in Figure 6.5B, we can explain some of the protection seen in the zinc-bound PMTR-DNA complex. Because the BmrR-DNA complex has two-fold symmetry, we need only inspect the PMTR-DNA



model for one PMTR monomer binding one DNA half-site. The following discussion will be framed around interactions with the left half-site of DNA shown in Figure 6.5B.

First, we see that the three bases near the center of the dyad are likely protected from cleavage on both strands by the distortion of the DNA itself and also by contacts from wing 2 of the winged helix motif, which is actually a second helix-turn-helix motif in BmrR. Bases -5 to -9 receive moderate to no protection, but are found near the PMTR recognition helix of the helix-turn-helix motif. Bases -10 to -11 are contacted by the wing 1 motif. The BmrR-DNA-drug structure was solved with a 22 base-pair oligonucleotide, so no contact information beyond bp 11 is available. It is very plausible that wing 1 might contact two more bases on the red strand, which may account for the observed hydroxyl radical footprint reaching to -13. The lavender strand approaches the protein most closely at bases 6 to 9, which agrees well with the 5 to 8 and hydroxyl radical footprints observed. The region of DNA protected by PMTR in footprinting studies aligns well with that protected by BmrR in similar studies (Ahmed et al., 1994), sharing 36% identity. Interestingly, the nucleotides that were hypersensitive to DNaseI cleavage in those studies also align well with those observed to be hypersensitive with PMTR.

Together, these data support the hypothesis that PMTR is a MerR family member that uses a DNA distortion mechanism similar to those of other MerR family members to either activate or repress transcription depending on the presence of an effector. To complete our structural understanding of the DNA binding properties of the MerR family, an apo protein-DNA structure is necessary to clarify the mechanism of repression by this family. We do see a decrease in protection of the DNA near the dyad axis in the absence

of zinc, which suggests that either the PMTR-DNA interface, the conformation of the DNA, or both has changed.

### **Acknowledgments**

The metal specificity assay was done by Svetlana Lutsenko. James Lundblad helped with EMSA and footprinting experiments, which I did in his laboratory. I carried out all other experiments reported in this chapter.

Table 6.1. Percent identity between operators bound by select MerR family members.

Operator bound by:	PMTR	CueR	ZntR	MerR	SoxR
CueR	75				
ZntR	15	24			
MerR	18	21	45		
SoxR	48	21	21	21	
BmrR	36	57	24	21	27

Table 6.2. Percent identity between select MerR family members, in total, in the DNA binding domain alone, and in the linker/effector binding domain. PM is PMTR, Zn is ZntR, Me is MerR, Cu is CueR, Pb is PbrR, Bm is BmrR, and Mt is Mta. BmrR and Mta were truncated at positions shown in Figure 6.8B.

	Total % identity for res. 1-150						% identity in DNA binding domain, res. 1-69						% identity in linker and effector binding domains, res. 70-150					
	PM	Zn	Me	Cu	Pb	Bm	PM	Zn	Me	Cu	Pb	Bm	Pm	Zn	Me	Cu	Pb	Bm
Zn	34						42						24					
Me	33	28					40	42					26	13				
Cu	37	34	32				35	33	34				39	28	31			
Pb	25	36	29	30			36	44	38	33			13	25	19	27		
Bm	19	19	19	18	15		28	31	26	30	19		7	7	11	6	9	
Mt	24	22	25	22	19	22	33	40	39	36	33	36	13	9	9	7	7	7

Figure 6.1A. Schematic of the *P. mirabilis* genomic fragment inserted into the PROT1 plasmid, containing the *hmpA*, *hmpB*, *atpase*, and *pmtr* genes. B. The intergenic regions, denoted *atpaseP* or *pmtrP*, used in EMSA and footprinting experiments. Transcription start sites are marked with arrows. Primers are drawn as half-arrows.

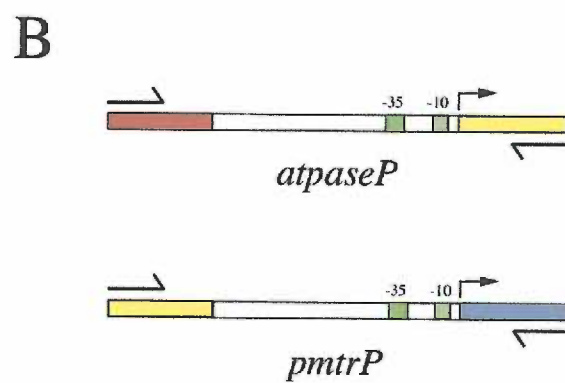
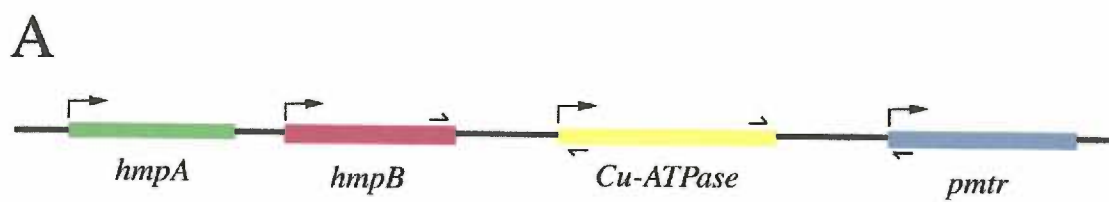
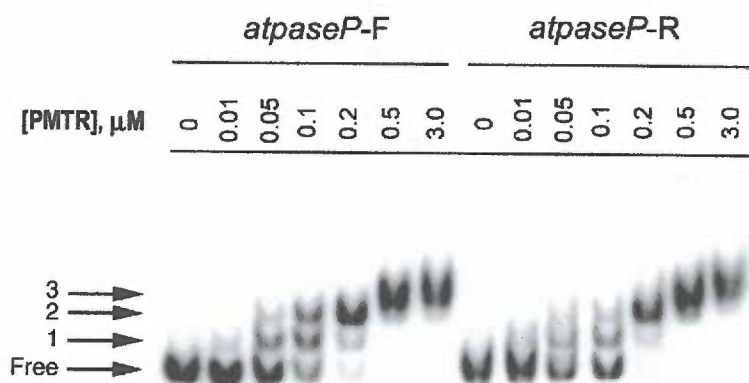


Figure 6.2. Electrophoretic mobility shift assay results for increasing concentrations of PMTR binding to *atpaseP* and *pmtrP*. Both radiolabelled strands (denoted "F" or "R") were tested for each promoter: A. *atpaseP* and B. *pmtrP*. Free DNA and higher order PMTR-DNA complexes are labelled.

A



B

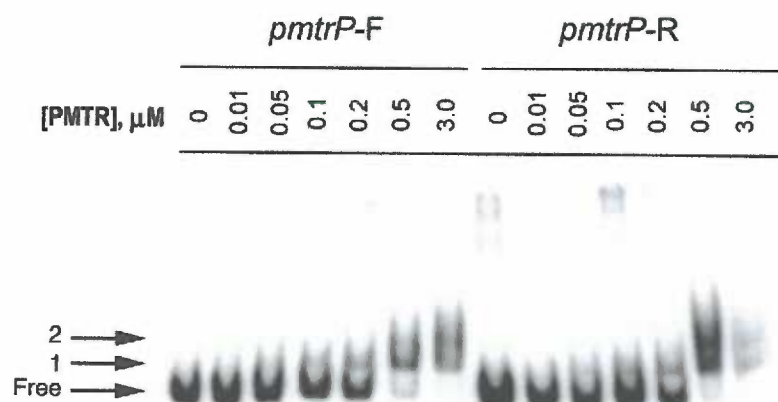




Figure 6.3. DNA cleavage protection experiments in the presence and absence of PMTR and 250  $\mu\text{M}$   $\text{ZnCl}_2$  with: A. *atpaseP-F* (labelled template strand) and B. *atpaseP-R* DNA (labelled non-template strand). Sequencing reactions are shown in lanes 1-4 and 15-18. DMS methylation protection reactions are shown in lanes 5-6 and 19-20. DNaseI interference reactions are shown in lanes 7-10 and 21-24. Hydroxyl radical interference cleavage reactions are in lanes 11-14 and 25-28. Reactions containing PMTR protein and/or 250  $\mu\text{M}$   $\text{ZnCl}_2$  are indicated with + or - above individual lanes. -35 and -10 promoter elements are marked with black lines and labelled. Relative positions to the putative *atpase* transcription start site are labelled with italic numbers, to the left of promoter element numbers.

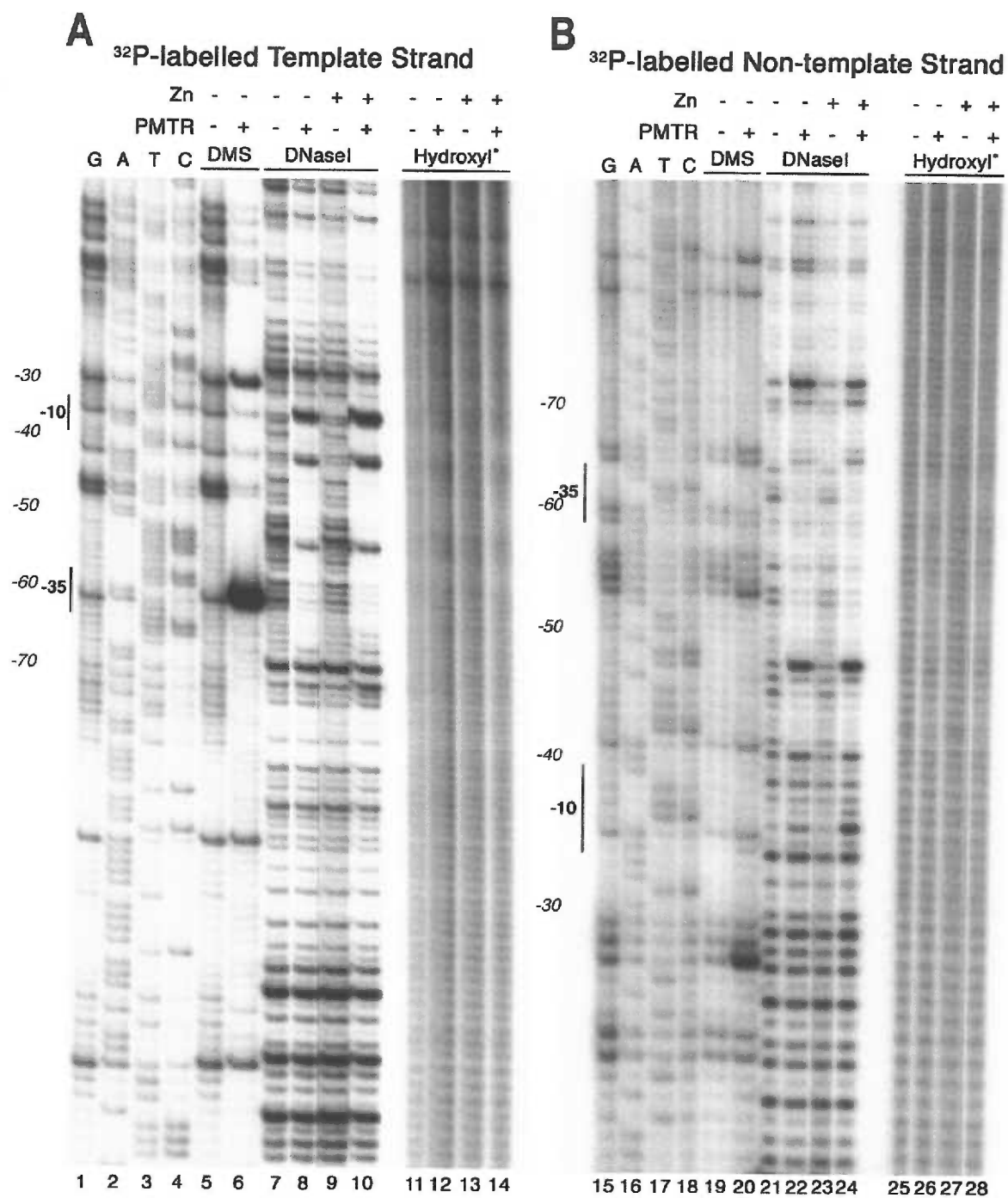
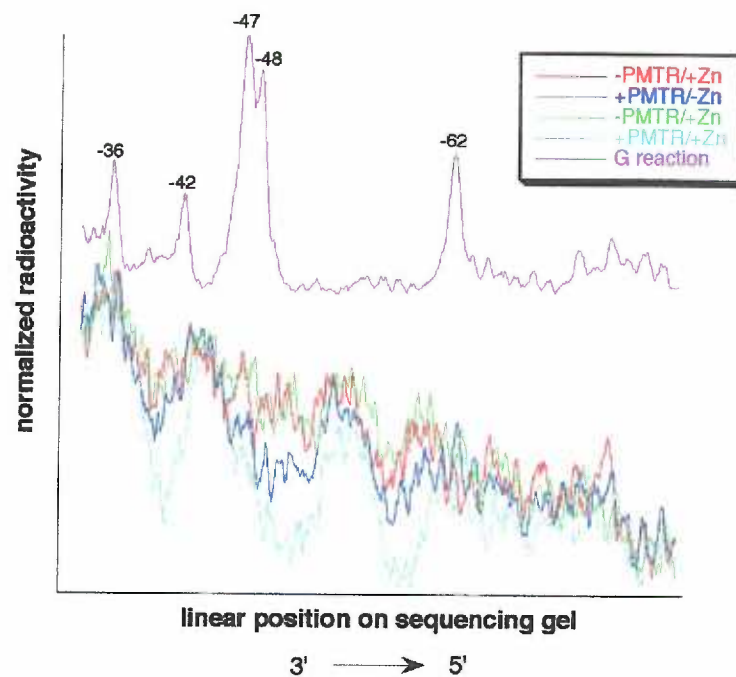


Figure 6.4. Linear traces of radioactivity down sequencing gel lanes from hydroxyl radical DNA cleavage reactions with *atpaseP* in the presence and absence of PMTR and 250 $\mu$ M ZnCl<sub>2</sub>. A. Template strand, *atpaseP*-F. B. Non-template strand, *atpaseP*-R. Lines are colored as listed in the legend. The G sequencing reaction linear trace is included for orientation within the promoter and is labelled with positions relative to the transcription start site.

## A Template strand hydroxyl radical cleavage



## B Non-template strand hydroxyl radical cleavage

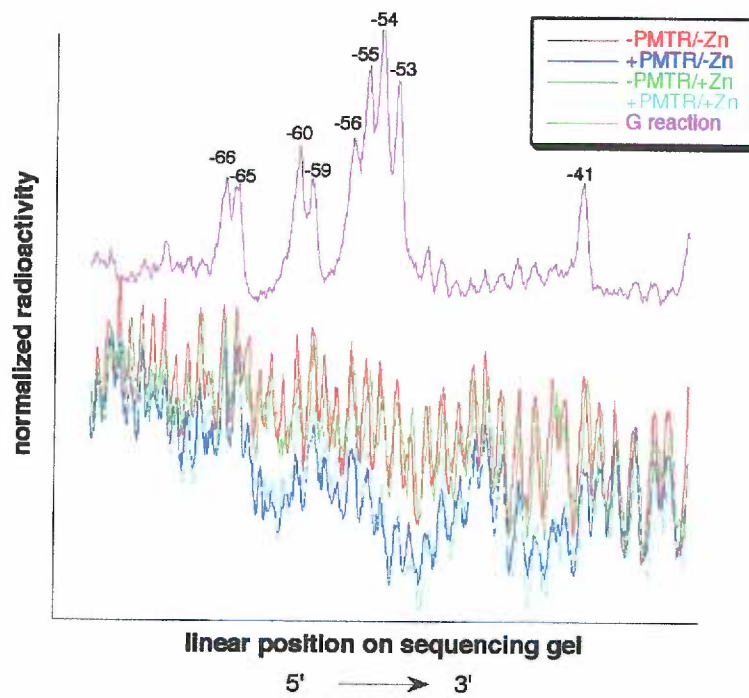
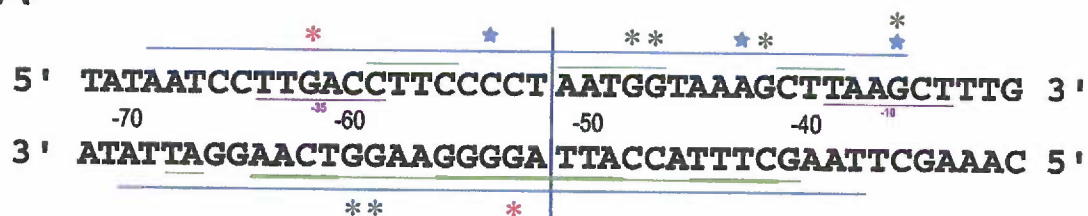
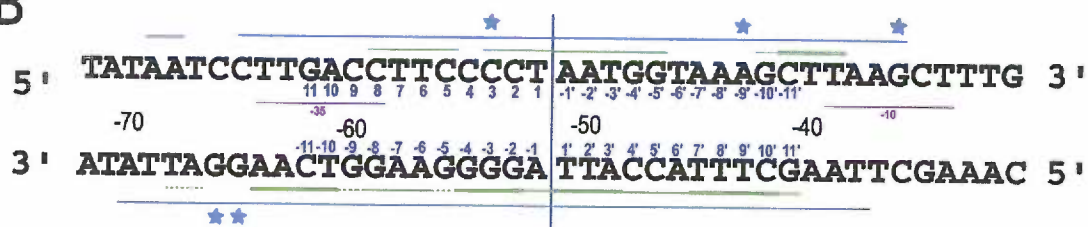


Figure 6.5. Summary of PMTR-*atpaseP* footprinting analyses. A. Results in the absence of zinc. B. Results in the presence of 250  $\mu\text{M}$   $\text{ZnCl}_2$ . In black are the positions of bases relative to the assumed translation start site, and in blue is the numbering scheme used in the BmrR-drug-DNA structure.

A



B



- Hydroxyl radical protection
- DNaseI protection
- ★ DNaseI hypersensitive
- \* DMS hypersensitive
- \* DMS protected

Figure 6.6. Fluorescence polarization equilibrium binding isotherms for PMTR binding to: A. the *atpaseP* operator identified in this study in the absence of zinc, B. the same operator in the presence of 250  $\mu$ M ZnCl<sub>2</sub>, C. the *E. coli zntA* operator, and D. the *E. coli purF* operator.

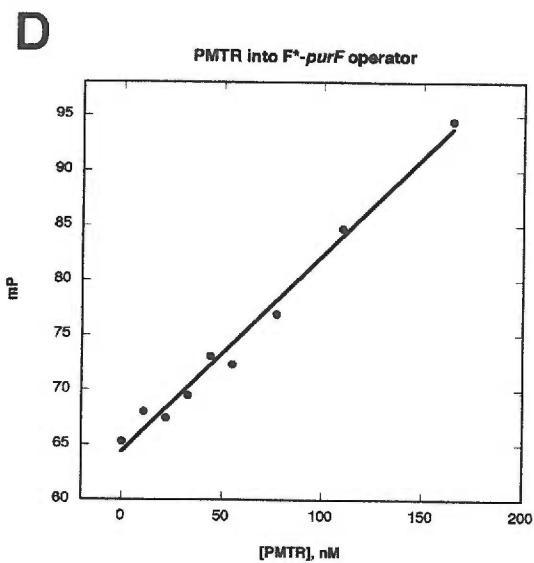
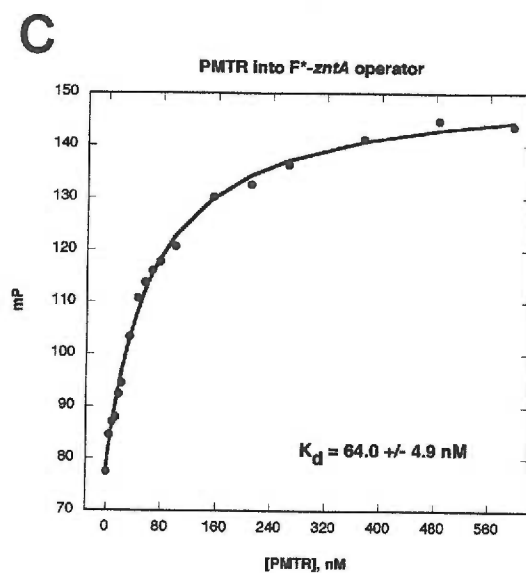
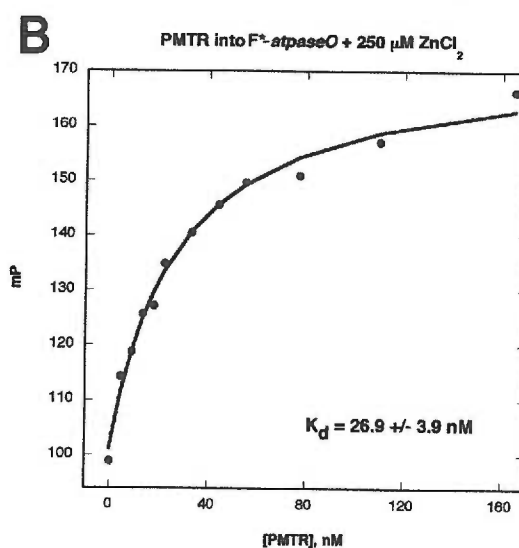
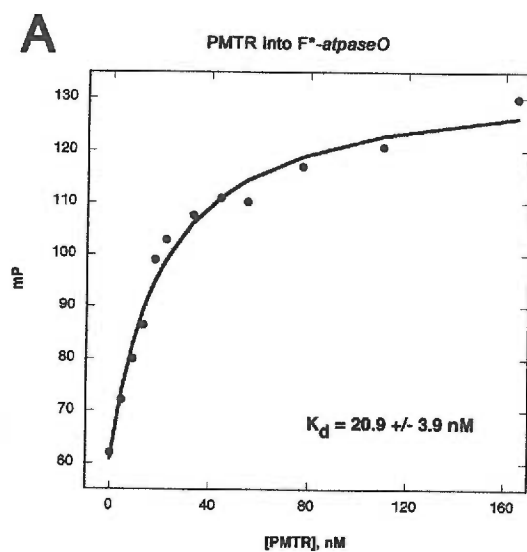
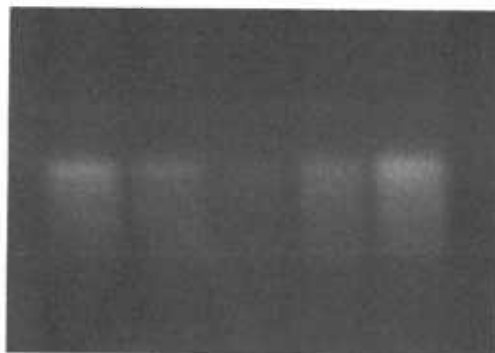


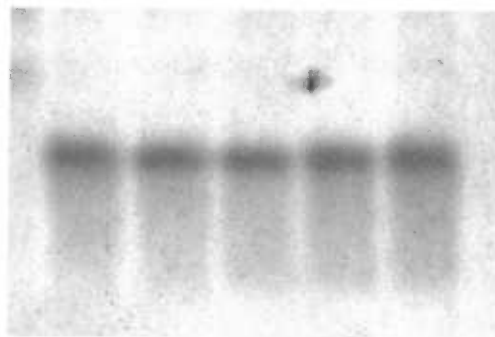


Figure 6.7. CPM labelling experiment results for PMTR in the presence or absence of copper, zinc, cadmium, or cobalt. A. CPM-labelled PMTR was visualized by UV light. B. The same gel stained with Coomassie Blue-containing dye and observed with visible light. C. Quantification of CPM labelling calculated as percent of labelling observed in the absence of metal.

A            -   Cu   Zn   Cd   Co



B



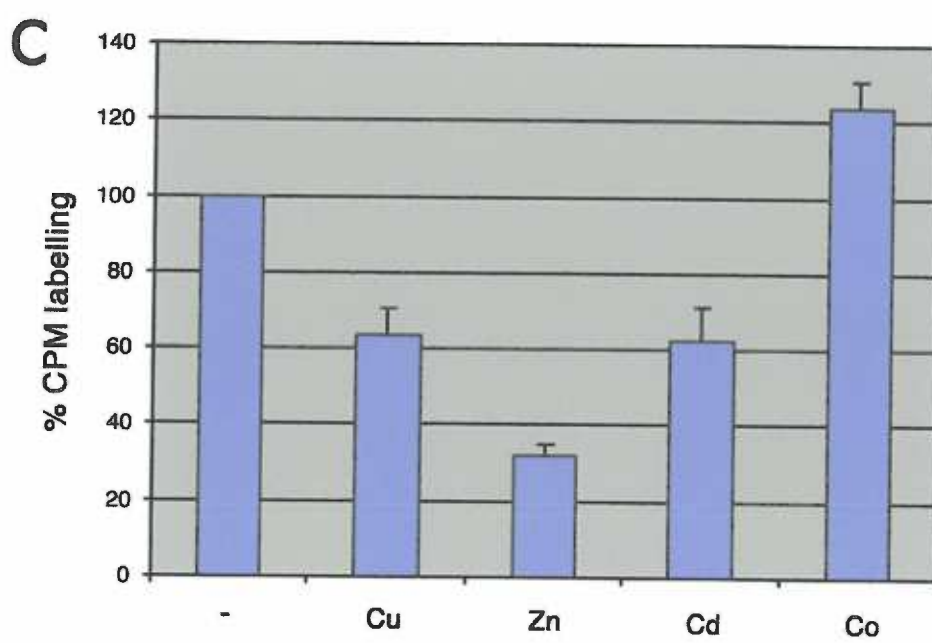


Figure 6.8. Sequence alignments of: A. the operator identified in our PMTR footprinting experiments aligned with operators bound by other MerR family members, and B. PMTR and select other MerR family members. BmrR and Mta were truncated to allow better sequence comparison.

## A Promoters bound by MerR family members

	-35	-10
PMTR_Pmirabilis	CTTGACCTTC	CCCTAATGGTAAAGCTTAAGCTT
CueR_Ecoli	CTTGACCTTC	CCCTTGCTGGGAAGGTTTAACCTT
ZntR_Ecoli	TTTGACTCTGGAGTCGA	CTCCAGAGTGTATCCTT
MerR_Ecoli	CTTGACTCCGTACATGAGTACGGAA	GTAAGGTT
SoxR_Ecoli	CTTTACCTCA	AGTTAACTTGAGGAATTATACTC
BmrR_Bsubtilis	GTTGACTCTC	CCCTAGGAGGAGGTCTTACAGTA

B

PMTR\_Proteus  
Zutr\_Escherichia  
Zutr\_Haemophilus  
MeRr\_Bacillus  
CuEr\_Escherichia  
PbRr\_Ralstonia  
BmRr\_Bacillus  
Ma\_Streptococcus

```

10 20 30 40 50 60
- - - - - MN I Q A A K N P V F S S K M I R Y Y E Q I G I I P K A - I R T D G G Y R D Y N D S D V D C F R I S H S R
- - - - - M Y R I G E L A N M A E V T P D I I K Y Y E K Q O O M E H E - V R T E G G F R L Y T E S D L Q R T K F I R H A R
- - - - - M K F R G E L A K A L G C T V E I I R Y Y E Q Q G L I P E P - K R T S G N F R Q Y N E E H Q R I S F I C N C R
- - - - - M K F R G E L A N I T G L T S K A I R F Y E E K G L V T P P - M R S E N G Y R M Y T Q Q H N E I T L L R Q A R
- - - - - M N I Q I G E L A K R T A C P V V I R F Y E Q E G L L P P P - G R S R G N F R L Y T G E H V E R I Q E T R H C R
- - - - - M K E S Y S I G E L A K L A N V S I K A L R Y Y D K I D I F K P A Y V D P D S Y R N Y T D S Q I I H L D L K S L K
- - - - - M Y H I K E A A Q L S G V S V K L H H D K I G L L V P . - L K S E N G Y R T Y S Q E D E R L Q V L L Y Y K

```

Recognition helix

Wing 1

PMTR\_Proteus  
Zutr\_Escherichia  
Zutr\_Haemophilus  
MeRr\_Bacillus  
CuEr\_Escherichia  
PbRr\_Ralstonia  
BmRr\_Bacillus  
Ma\_Streptococcus

```

70 80 90 100 110 120
A L G F S T E Q I S T L I V L W N N R E R T - S A D V K A I A L S H I D E L N R K T Q L Q R M T Q - - - - T S H L
Q L G F S L E S I R E I S I R I D P E H H T C O E S K G I V Q E R L Q E V E A R A E I Q S M Q R - - - - S L Q R L
N I D I S I S E I K S I N L E N A S K Q Q - A E E I N R V L D K H I K E V A T R I H E I A H L R M - - - - K I L E L
E L G F T N E I D K L G V V D R D E A K - C R D V M Y D F T I L K I E D I Q R K I E D L O S M R D - - - - Q L L A L
Q V G F N L E E S G E I V N L F N D P Q R H - S A D V K R R P D Q D - C G E V N M L L D E H I R Q V E S R G A L L E L K H - - - - H L V E
S I D M P L S D V R T I S Y R K R P D Q D - C G E V N M L L D E H I R Q V E S R G A L L E L K H - - - - H L V E
Y I G T P L E E M K K A O D L E M E L F A F Y T E Q E R Q I R E K L D F L S A L E Q I S L V K K R - - - - M K R Q M E Y
Y L G F S L E K I A E L K E E R T D L L P H L T R Q L D Y L T R E R Q H L D T L S T Q K T I Q E Q K G E R K M T I

```

Wing 2 - Second HTH

Linker region

PMTR\_Proteus  
Zutr\_Escherichia  
Zutr\_Haemophilus  
MeRr\_Bacillus  
CuEr\_Escherichia  
PbRr\_Ralstonia  
BmRr\_Bacillus  
Ma\_Streptococcus

```

130 140 150
A Q E C Q Q G - - - D N N P D C P I I A K I V E P Q T G T E H - - -
N D A C C G G T A H S S V Y C S I L E A L E Q G A S G V K S G C -
R E R C T V S - N D E D P M K L L L Q H S G V K F V R L K - - -
K E R C P E - N K D I Y E C P T I E T M K K - - - - -
A N A C P G - - D D S A D C P I I E N S G C C H R A G - - -
R E A C S G - A R P A O S C G I L O G L S D C V C D T R G T A
P A L G E V F V L D E E I R I I Q T E A B G I G P E N V L N A
E E K F T G F S Y Q D N Q K Y H Q E A V E K Y G Q E V M G Q A L

```

Effector binding domain

Figure 6.9. Model of the dimeric PMTR-DNA complex. The model was constructed using the BmrR-TPP-DNA complex (Heldwein and Brennan, 2001) as a template for the PMTR sequence. The DNA sequence was not mutated. The helix-turn-helix motif is labelled HTH, wing 1 is labelled W1, and wing 2 is labelled W2. Residues corresponding to metal-binding residues of MerR are shown.

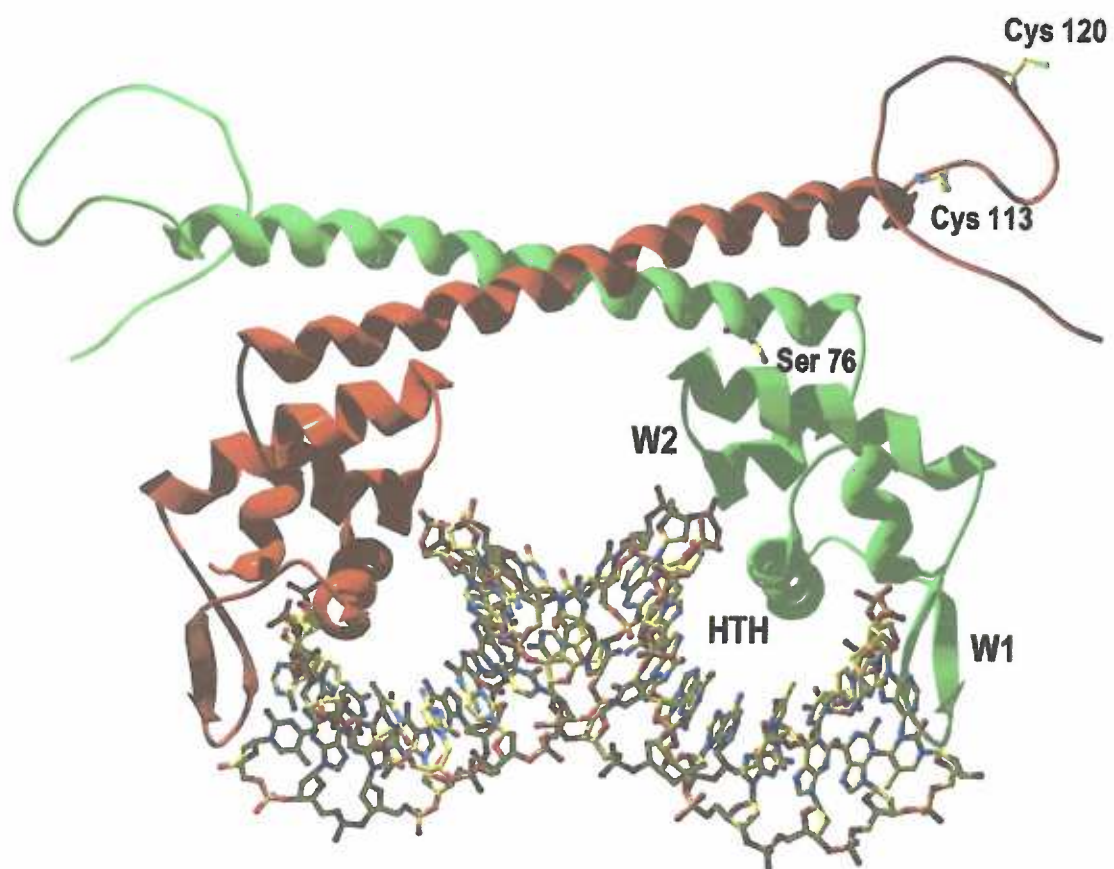
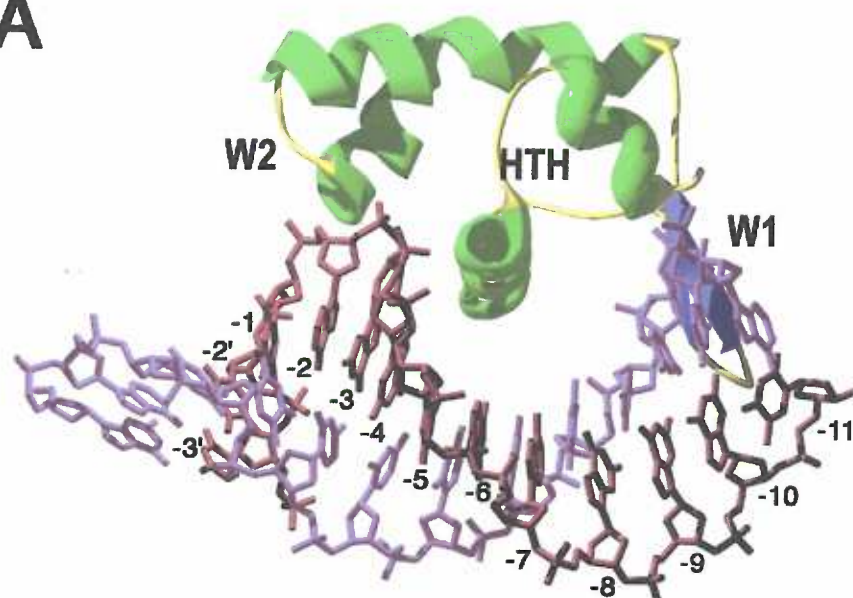


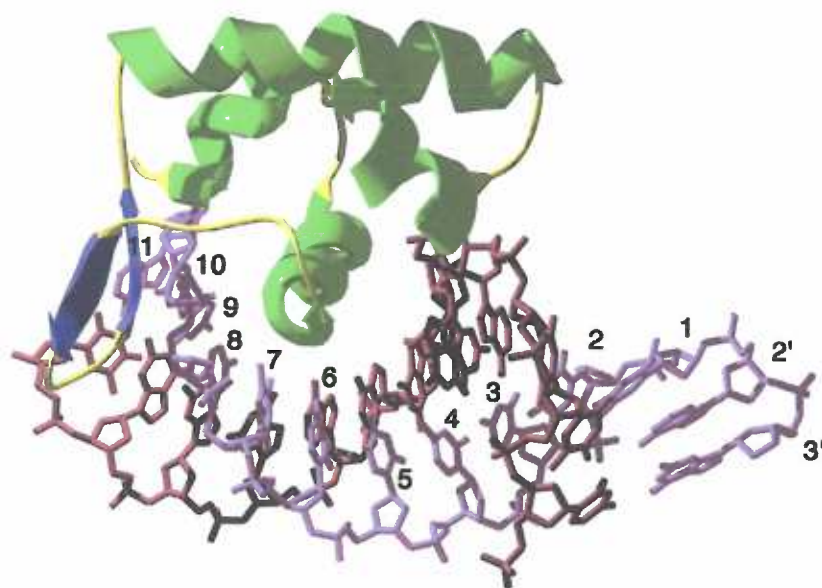


Figure 6.10. Close-up view of the model of one PMTR DNA binding domain monomer bound to one DNA half-site. The lavender strand represents the positively numbered bases in Figure 6.5B, and the red strand corresponds to the negatively numbered bases. A. A similar orientation to that seen in Figure 6.9, but closer to the green monomer. B. The same monomer, rotated 180° about the conventional Z axis.

**A**



**B**



## Summary and conclusions

The work presented in this thesis examines two aspects of transcription regulation in prokaryotes. The first relates to corepressor binding and the allosteric switch within the *E. coli* purine repressor, PurR. The second is the biochemical characterization of the *P. mirabilis* transcription regulator, PMTR. In both cases, the studies involve prokaryotic transcription regulators, but are at quite different levels and employ overlapping and different tools due to the abundance of biochemical, genetic, and structural data available for the former and the dearth of information about the latter.

Prior to this work, much biochemical and genetic information had been gathered about PurR (reviewed in Chapter 1 and (Zalkin and Nygaard, 1996)), and the crystal structures of PurR in its DNA and corepressor bound and apo conformations had been solved (Schumacher et al., 1995; Schumacher et al., 1994a). This was an excellent starting place for the structural and biochemical studies described in Chapters 3-5 of this thesis, which involved understanding the function of particular residues in corepressor binding and the allosteric switch mechanism.

Site-directed mutagenesis and structural data showed residue Trp147 as important in the initiation of the intramolecular signal transduction—or allosteric switch—from the corepressor binding domain to the DNA binding domain that corepressor had bound (Lu et al., 1998a; Schumacher et al., 1995). In Chapter 3, I described the structural and biochemical analyses of three substitutions at position 147. These studies highlighted the role a single residue can play in regulating global conformational changes. Substitution of residue Trp147 with Phe resulted in a weaker repressor, while mutation to Ala or Arg resulted in "super" repressors *in vivo*. The *in vitro* corepressor binding data correlated

well with the *in vivo* results, as corepressor affinities were higher for super repressors and lower for the weaker repressor. DNA binding studies showed that the altered repression activities did not stem from the inability of the mutant PurR proteins to bind DNA.

The crystal structures of each of the mutants complexed with hypoxanthine and *purF* operator suggested reasons for the altered ligand binding and hence *in vivo* activities of the mutants. To explain the changes in affinity of each of the mutant proteins for corepressor, each must either favor one conformational end state or disfavor the other, or both. W147F was a weaker repressor because it made more favorable interactions in the apo "open" conformation than in the corepressor-bound "closed" state. W147A was a super repressor because the open conformation was destabilized by the lack of contacts made in the corepressor binding pocket, while in the closed conformation, the allosteric loop was altered somewhat to fill the cavity created by the removal of the indole side chain. W147R was a super repressor because its long, charged side chain could not be accommodated in the corepressor binding pocket that is occupied by the tryptophan in the open conformation. Additionally, the location of the arginine in the ligand bound form is well accommodated by the indole binding pocket.

PurR, LacI, and TreR are the only members of the LacI family for which structures of the corepressor binding domain are available. Neither LacI nor TreR has a residue analogous to Trp147, as judged from a structure-based alignment of their sequences (Hars et al., 1998). In LacI and TreR, the residues at corresponding positions to Trp147 contact their small molecule pyranose inducers. This difference might be ascribed to the mechanism of PurR by which it represses in the presence of its ligand, which is a product of the anabolic pathway it regulates. On the other hand, most other

LacI family members are induced by ligands, which are substrates of the catabolic pathways that they regulate. Interestingly, RbsR, the evolutionarily closest relative of PurR, has a tryptophan at the position equivalent to PurR Trp147 yet is induced by ribose. Structural information regarding the function of this residue in RbsR might clarify the role of this residue in PurR as it relates to other LacI family members and provide a glimpse of evolution.

Substitution of other residues involved in the allosteric switch within PurR might elucidate the relative importance of Trp147 in effecting the intramolecular signal. At the interface of the two monomers lie several residues, including Glu70 and Tyr107, which form hydrogen bonds across the symmetry axis that are broken upon binding of corepressor. The mutation Glu70Ala has been made and shown to have increased repression activity over wild-type, as was seen with Trp147Ala (Lu et al., 1998a).

Mutation of other residues in the "switch" loop on which Trp147 resides, encompassing residues 146-154, might reveal whether flexibility is required or whether other residues play a part in the motion of the loop. Among these, Asp146, Gly148, and Lys151 might prove most revealing. The first two are proximal to Trp147 on the polypeptide chain, and the Lys151 side chain is near Trp147 in the holo complex structure. Mutagenesis studies have previously shown that an Asp146Ala substitution reduced *in vivo* activity, but structural analysis of this mutant has not yet been performed (Choi et al., 1994). Mutation of other residues found in the apo corepressor binding pocket, such as Tyr73, Asp160, or Asn161, might illustrate the roles of these side chains, and particularly the importance of the hydrogen bond between them, in the apo conformer allosteric switch.

The studies reported in Chapter 4 on PurR mutant proteins, substituted at position Glu222, have confirmed the role of this residue in determining the specificity of PurR for 2-aminopurines. Such a role for the Glu222 side chain was posited in earlier work (Schumacher et al., 1997), but the experiments described in this thesis definitively established the relationship between the hydrogen bonding and electrostatic complementarity of the side chain of residue 222 and the identity of the purine corepressor. Moreover, this work revealed the importance of the electrostatics of the entire corepressor binding pocket with respect to corepressor specificity. In wild-type PurR, the carboxylate group of Glu222 and the carbonyl oxygen of Phe221 accept hydrogen bonds from a water molecule, thus leaving the water molecule with only the ability to accept hydrogen bonds from any other hydrogen bonding partners and resulting in a water-specific interaction with the hydrogen bond donating N2 nitrogen atom of guanine.

The E222A and E222Q mutant proteins had reduced affinity for guanine and hypoxanthine, with a slight loss of preference for guanine over hypoxanthine, while affinity for xanthine was increased. In the E222A mutant, this was due to the loss of the hydrogen bonding capability of the alanine side chain and a slightly improved electrostatic complementarity of the binding pocket to xanthine. In the E222Q mutant, the glutamine side chain is electrostatically neutral, like the alanine side chain, but retains the ability to hydrogen bond to corepressor similar to the wild-type glutamate. Surprisingly, the glutamine side chain does not make the same water-mediated hydrogen bond to the exocyclic 2 position of the purine as found in the wild-type guanine complex. Instead, the glutamine side chain is bent away from the binding pocket, which accounts

for its comparable corepressor affinities with E222A. As predicted, the Glu222Lys mutation confers higher affinity for xanthine, with a concurrent loss of affinity for the physiological corepressors, hypoxanthine and guanine. This is the direct result of the increased electrostatic complementarity between the Nε group of lysine and the xanthine O2 and the retention of the water-mediated hydrogen bond to the O2. Mutation of the negatively charged residues near Glu222 in the corepressor binding pocket, such as Glu220 and Glu70, could reveal more about the electrostatic contribution to corepressor specificity.

In Chapter 5, binding data were presented for mutant proteins R190K, T192P, and R190K/T192P, which were designed to redesign corepressor specificity in PurR from hypoxanthine and guanine to ribose. A slight improvement in affinity was observed for R190K, concomitant with a slight decrease in affinity for hypoxanthine. Unfortunately, the Thr192Pro mutation destabilized the protein and prohibited its investigation. Although this simplistic approach failed to produce a ribose-regulated PurR, several options remain to succeed in this endeavor. The first is to mutate the entire loop, residues 189-195, on which Arg190 and Thr192 reside, to the corresponding RbsR sequence. Another possibility is switching the positions of Tyr73 and Phe74, which are Phe73 and Tyr74 in RbsR. These are the only other residues that contact corepressor in PurR that differ between the two proteins. Perhaps additional residues are used by RbsR to bind ribose. It is also possible that more subtlety is required for the evolution of specificity than we initially appreciated. We may not know all we need to about ligand specificity, and our failed attempts may serve as a warning against drawing conclusions from modelling studies. Furthermore, this work may be an admonition against the idea

proposed by structural proteomics that a few representative structures from a particular family are enough to predict molecular interactions at the atomic level for other family members.

The second major topic of this thesis focused on the initial biochemical characterization of the *Proteus mirabilis* transcription factor, PMTR, a MerR family member. To my knowledge, this represents the first detailed analysis of any *P. mirabilis* transcription factor. Furthermore, PMTR is one of only six metal-binding MerR family members to have its DNA binding site identified. This work includes an experiment that characterizes the metal binding affinity of PMTR *in vitro*, which represents the first *in vitro* metal specificity determination for any MerR family member. The specificity appears to be different from either *E. coli* ZntR or CueR, as determined *in vivo* (Binet and Poole, 2000; Stoyanov et al., 2001), in that PMTR binds zinc, cadmium, and copper. Many experiments are possible now that the metal and DNA specificities have been determined. The first is to confirm, *in vivo*, that the promoter identified in Chapter 6 is activated by PMTR in response to increased metal levels. An *in vivo* reporter assay is currently being developed to test this, and metal specificity can be confirmed using the same construct. The metal specificity experiments suggested that PMTR uses cysteines to bind metal, and PMTR has two cysteines that correspond to two of the three cysteines used by MerR to bind mercury, based on sequence alignments. It is tempting to speculate that Cys112 and Cys120 of PMTR are involved in its metal binding. These putative metal binding residues could be mutated and the effects of the mutations determined with the *in vivo* repression assay. To identify other residues involved in metal coordination, random mutagenesis of the effector binding domain could be employed. Of course, the



future goals for this project are crystallization and structure determination, although the limited solubility of the protein must still be overcome. Perhaps mutating residues prone to oxidation, such as cysteines, or mutating surface-exposed hydrophobic residues would increase the solubility of the protein. The activity of such a mutant might be altered, a fact that would have to be weighed carefully. Alternatively, if a homologous protein from another organism were found, structural studies might be possible with that protein.

## Literature cited

- Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution, *Science* 242, 899-907.
- Ahmed, M., Borsch, C. M., Taylor, S. S., Vasquez-Laslop, N., and Neyfakh, A. A. (1994). A protein that activates expression of a multidrug efflux transporter upon binding the transporter substrates, *J. Biol. Chem.* 269, 28506-28513.
- Ahmed, M., Lyass, L., Markham, P. N., Taylor, S. S., Vasquez-Laslop, N., and Neyfakh, A. A. (1995). Two highly similar multidrug transporters of *Bascillus subtilis* whose expression is differentially regulated, *J. Bacteriol.* 177, 3904-3910.
- Alberts, I. L., Nadassy, K., and Wodak, S. J. (1998). Analysis of zinc binding sites in protein crystal structures, *Protein Sci.* 7, 1700-1716.
- Albright, R. A., and Matthews, B. W. (1998a). Crystal structure of lambda-Cro bound to a consensus operator at 3.0 Å resolution, *J. Mol. Biol.* 280, 137-51.
- Albright, R. A., and Matthews, B. W. (1998b). How Cro and lambda-repressor distinguish between operators: the structural basis underlying a genetic switch, *Proc. Natl. Acad. Sci. U.S.A.* 95, 3431-6.
- Anderson, W. F., Ohlendorf, D. H., Takeda, Y., and Matthews, B. W. (1981). Structure of the cro repressor from bacteriophage lambda and its interaction with DNA, *Nature* 290, 754-8.
- Ansari, A. Z., Bradner, J. E., and O'Halloran, T. V. (1995). DNA-bend modulation in a repressor-to-activator switching mechanism, *Nature* 374, 371-5.

- Ansari, A. Z., Chael, M. L., and O'Halloran, T. V. (1992). Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR, *Nature* 355, 87-9.
- Arvidson, D. N., Lu, F., Faber, C., Zalkin, H., and Brennan, R. G. (1998). The structure of PurR mutant L54M shows an alternative route to DNA kinking, *Nat. Struct. Biol.* 5, 436-41.
- Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R. P., and Dickerson, R. E. (1996). Structure of the *Escherichia coli* response regulator NarL, *Biochemistry* 35, 11053-61.
- Baranova, N. N., Danchin, A., and Neyfakh, A. A. (1999). Mta, a global MerR-type regulator of the *Bacillus subtilis* multidrug-efflux transporters, *Mol. Microbiol.* 31, 1549-1559.
- Bell, C. E., and Lewis, M. (2000). A closer view of the conformation of the Lac repressor bound to operator [see comments], *Nat. Struct. Biol.* 7, 209-14.
- Binet, M. R., and Poole, R. K. (2000). Cd(II), Pb(II) and Zn(II) ions regulate expression of the metal-transporting P-type ATPase ZntA in *Escherichia coli*, *FEBS Lett.* 473, 67-70.
- Blundell, T. L., and Johnson, L. N. (1976). Protein crystallography, 1st edn (San Diego, California, Academic Press, Inc.).
- Bragg, W. L. (1913). The diffraction of short electromagnetic waves by a crystal, *Proc. Camb. Phil. Soc.* 17, 43-57.
- Breg, J. N., van Opheusden, J. H., Burgering, M. J., Boelens, R., and Kaptein, R. (1990). Structure of Arc repressor in solution: evidence for a family of  $\beta$ -sheet DNA-binding proteins, *Nature* 346, 586-9.

- Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff, *Cell* 74, 773-6.
- Brennan, R. G., and Matthews, B. W. (1989). The helix-turn-helix DNA binding motif, *J. Biol. Chem.* 264, 1903-6.
- Brennan, R. G., Roderick, S. L., Takeda, Y., and Matthews, B. W. (1990). Protein-DNA conformational changes in the crystal structure of a lambda Cro-operator complex, *Proc. Natl. Acad. Sci. U.S.A.* 87, 8165-9.
- Brocklehurst, K. R., Hobman, J. L., Lawley, B., Blank, L., Marshall, S. J., Brown, N. L., and Morby, A. P. (1999). ZntR is a Zn(II)-responsive MerR-like transcriptional regulator of *zntA* in *Escherichia coli*, *Mol. Microbiol.* 31, 893-902.
- Brown, N. L., Rouch, D. A., and Lee, B. T. O. (1992). Copper resistance determinants in bacteria, *Plasmid* 27, 41-51.
- Bruins, M. R., Kapil, S., and Oehme, F. W. (2000). Microbial resistance to metals in the environment, *Ecotoxicol. Environ. Saf.* 45, 198-207.
- Burgering, M. J., Boelens, R., Gilbert, D. E., Breg, J. N., Knight, K. L., Sauer, R. T., and Kaptein, R. (1994). Solution structure of dimeric Mnt repressor (1-76), *Biochemistry* 33, 15036-45.
- Burley, S. K., and Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization, *Science* 229, 23-28.
- Busby, S., and Ebright, R. H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes, *Cell* 79, 743-6.
- Busby, S., and Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP), *J. Mol. Biol.* 293, 199-213.

- Campos, A., Zhang, R., Alkire, R. W., Matsumura, P., and Westbrook, E. M. (2001). Crystal structure of the global regulator FlhD from *Escherichia coli* at 1.8 Å resolution, *Mol. Microbiol.* 39, 567-580.
- Chen, Y., and Rosen, B. P. (1997). Metalloregulatory properties of the ArsD repressor, *J. Biol. Chem.* 272, 14257-62.
- Choi, K. Y., Lu, F., and Zalkin, H. (1994). Mutagenesis of amino acid residues required for binding of corepressors to the purine repressor, *J. Biol. Chem.* 269, 24066-72.
- Choi, K. Y., and Zalkin, H. (1990). Regulation of *Escherichia coli* pyrC by the purine regulon repressor protein, *J. Bacteriol.* 172, 3201-7.
- Choi, K. Y., and Zalkin, H. (1992). Structural characterization and corepressor binding of the *Escherichia coli* purine repressor, *J. Bacteriol.* 174, 6207-14.
- Choi, K. Y., and Zalkin, H. (1994). Role of the purine repressor hinge sequence in repressor function, *J. Bacteriol.* 176, 1767-72.
- Clackson, T. (1998). Redesigning small molecule-protein interfaces, *Curr. Opin. Struct. Biol.* 8, 451-8.
- Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA recognition motif resembles histone H5, *Nature* 364, 412-420.
- Clubb, R. T., Omichinski, J. G., Savilahti, H., Mizuuchi, K., Gronenborn, A. M., and Clore, G. M. (1994). A novel class of winged helix-turn-helix protein: the DNA-binding domain of Mu transposase, *Structure* 2, 1041-8.
- Cohn, E. J., and Ferry, J. D. (1943). The interactions of proteins with ions and dipolar ions. In *Proteins, amino acids, and peptides as ions and dipolar ions*, E. J. Cohn, and J. T. Edsall, eds. (Princeton, New Jersey, Van Nostrand-Reinhold), pp. 586.

- Coleman, J. E. (1992). Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins, *Annu. Rev. Biochem.* 61, 897-946.
- Cook, W. J., Kar, S. R., Taylor, K. B., and Hall, L. M. (1998). Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins, *J. Mol. Biol.* 275, 337-46.
- de Champs, C., Bonnet, R., Sirot, D., Chanal, C., and Sirot, J. (2000). Clinical relevance of *Proteus mirabilis* in hospital patients: a two year survey, *J. Antimicrob. Chemother.* 45, 537-9.
- Declerck, N., and Abelson, J. (1994). Novel substrate specificity engineered in the arabinose binding protein, *Protein Eng.* 7, 997-1004.
- Delgado, J., Forst, S., Harlocker, S., and Inouye, M. (1993). Identification of a phosphorylation site and functional analysis of conserved aspartic acid residues of OmpR, a transcriptional activator for ompF and ompC in *Escherichia coli*, *Mol. Microbiol.* 10, 1037-47.
- Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality, *Nucleic Acids Res.* 26, 1906-26.
- Dixon, W. J., Hayes, J. J., Levin, J. R., Weidner, M. F., Dombroski, B. A., and Tullius, T. D. (1991). Hydroxyl radical footprinting, *Methods Enzymol.* 208, 380-413.
- Drenth, J. (1994). Principles of protein X-ray crystallography, 2nd edn (New York, Springer).
- Dworkin, J., Jovanovic, G., and Model, P. (1997). Role of upstream activation sequences and integration host factor in transcriptional activation by the constitutively active prokaryotic enhancer-binding protein PspF, *J. Mol. Biol.* 273, 377-88.

- Endo, G., and Silver, S. (1995). CadC, the transcriptional regulatory protein of the cadmium resistance system of *Staphylococcus aureus* plasmid pI258, *J. Bacteriol.* 177, 4437-41.
- Ewald, P. P. (1969). Introduction to the dynamical theory of X-ray diffraction, *Acta Crystallogr. A* 25, 103-108.
- Fitzgerald, P. M. D. (1988). MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement, *J. Appl. Crystallog.* 26, 283-291.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., and et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science* 269, 496-512.
- Fogh, R. H., Otteleben, G., Ruterjans, H., Schnarr, M., Boelens, R., and Kaptein, R. (1994). Solution structure of the LexA repressor DNA binding domain determined by <sup>1</sup>H NMR spectroscopy, *EMBO J.* 13, 3936-44.
- Frantz, B., and O'Halloran, T. V. (1990). DNA distortion accompanies transcriptional activation by the metal- responsive gene-regulatory protein MerR, *Biochemistry* 29, 4747-51.
- Friedman, A. M., Fischmann, T. O., and Steitz, T. A. (1995). Crystal structure of *lac* repressor core tetramer and its implications for DNA looping, *Science* 268, 1721-1727.
- Friedrich, W., Knipping, P., and Laue, M. (1912). Interferenz-Erscheinungen bei Roentgenstrahlen, tr."Interference phenomena for X-rays", *Bayerische Akad. Wiss.*, 303-322.

- Gajiwala, K. S., and Burley, S. K. (2000). Winged helix proteins, *Curr. Opin. Struct. Biol.* *10*, 110-6.
- Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B., and Burley, S. K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding, *Nature* *403*, 916-21.
- Gaudu, P., and Weiss, B. (1996). SoxR, a [2Fe-2S] transcription factor, is active only in its oxidized form, *Proc. Natl. Acad. Sci. U.S.A.* *93*, 10094-8.
- Glasfeld, A., Koehler, A. N., Schumacher, M. A., and Brennan, R. G. (1999). The role of lysine 55 in determining the specificity of the purine repressor for its operators through minor groove interactions, *J. Mol. Biol.* *291*, 347-61.
- Glasfeld, A., Schumacher, M. A., Choi, K.-Y., Zalkin, H., and Brennan, R. G. (1996). A positively charged residue bound in the minor groove does not alter the bending of a DNA duplex., *J. Amer. Chem. Soc.* *118*, 13073-13074.
- Gomis-Rüth, F. X., Sola, M., Acebo, P., Parraga, A., Guasch, A., Eritja, R., Gonzalez, A., Espinosa, M., del Solar, G., and Coll, M. (1998). The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator, *EMBO J.* *17*, 7404-7415.
- Gralla, J. D. (1996). Activation and repression of *E. coli* promoters, *Curr. Opin. Genet. Dev.* *6*, 526-30.
- Green, A. A. (1932). The solubility of hemoglobin in solutions of chlorides and sulfates of varying concentrations, *J. Biol. Chem.* *95*, 47.



- Grosse, C., Grass, G., Anton, A., Franke, S., Santos, A. N., Lawley, B., Brown, N. L., and Nies, D. H. (1999). Transcriptional organization of the *czc* heavy-metal homeostasis determinant from *Alcaligenes eutrophus*, *J. Bacteriol.* *181*, 2385-93.
- Haldimann, A., Prahalad, M. K., Fisher, S. L., Kim, S. K., Walsh, C. T., and Wanner, B. L. (1996). Altered recognition mutants of the response regulator PhoB: a new genetic strategy for studying protein-protein interactions, *Proc. Natl. Acad. Sci. U.S.A.* *93*, 14361-6.
- Hall, D. R., Gourley, D. G., Leonard, G. A., Duke, E. M., Anderson, L. A., Boxer, D. H., and Hunter, W. N. (1999). The high-resolution crystal structure of the molybdate-dependent transcriptional regulator (ModE) from *Escherichia coli*: a novel combination of domain folds, *EMBO J.* *18*, 1435-46.
- Harding, M. M. (1999). The geometry of metal-ligand interactions relevant to proteins, *Acta Crystallogr. D Biol. Crystallogr.* *D55*, 1432-1443.
- Hars, U., Horlacher, R., Boos, W., Welte, W., and Diederichs, K. (1998). Crystal structure of the effector-binding domain of the trehalose- repressor of *Escherichia coli*, a member of the LacI family, in its complexes with inducer trehalose-6-phosphate and noninducer trehalose, *Protein Sci.* *7*, 2511-21.
- He, B., Choi, K. Y., and Zalkin, H. (1993). Regulation of *Escherichia coli* *glnB*, *prsA*, and *speA* by the purine repressor, *J. Bacteriol.* *175*, 3598-606.
- He, B., Shiau, A., Choi, K. Y., Zalkin, H., and Smith, J. M. (1990). Genes of the *Escherichia coli* *pur* regulon are negatively controlled by a repressor-operator interaction, *J. Bacteriol.* *172*, 4555-62.

- He, B., Smith, J. M., and Zalkin, H. (1992). *Escherichia coli purB* gene: cloning, nucleotide sequence, and regulation by purR, J. Bacteriol. 174, 130-6.
- Hedstrom, L., Szilagyi, L., and Rutter, W. J. (1992). Science 255, 1249-53.
- Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (1992). Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target, Nature 359, 505-12.
- Heldwein, E. E., and Brennan, R. G. (2001). Crystal structure of the transcription activator BmrR bound to DNA and a drug, Nature 409, 378-82.
- Helmann, J. D., Ballard, B. T., and Walsh, C. T. (1990). The MerR regulatory protein binds mercuric ion as a tricoordinate, metal-bridged dimer, Science 247, 946-8.
- Herschbach, B. M., and Johnson, A. D. (1993). Transcriptional repression in eukaryotes, Annu. Rev. Cell Biol. 9, 479-509.
- Hidalgo, E., Ding, H., and Dimple, B. (1997). Redox signal transduction via iron-sulfur clusters in the SoxR transcription activator, Trends Biochem. Sci. 22, 207-10.
- Hidalgo, E., Leautaud, V., and Dimple, B. (1998). The redox-regulated SoxR protein acts from a single DNA site as a repressor and an allosteric activator, EMBO J. 17, 2629-2636.
- Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W., and Saenger, W. (1994). Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance, Science 264, 418-20.
- Holmes, D. J., Caso, J. L., and Thompson, C. J. (1993). Autogenous transcriptional activation of a thiostrepton-induced gene in *Streptomyces lividans*, EMBO J. 12, 3183-3191.

- Holmes, K. C., and Blow, D. M. (1966). The use of X-ray diffraction in the study of protein and nucleic acid structure (New York, Interscience).
- Ilangovan, U., Wojciak, J. M., Connolly, K. M., and Clubb, R. T. (1999). NMR structure and functional studies of the Mu repressor DNA-binding domain, *Biochemistry* 38, 8367-76.
- Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins, *J. Mol. Biol.* 3, 318-356.
- Jacquet, M.-A., and Reiss, C. (1990). Transcription *in vivo* directed by consensus sequences of *E. coli* promoters: their context heavily affects efficiencies and start sites, *Nucleic Acids Res.* 18, 1137-43.
- Jia, X., Grove, A., Ivancic, M., Hsu, V. L., Geiduscheck, E. P., and Kearns, D. R. (1996). Structure of the *Bacillus subtilis* phage SPO1-encoded type II DNA-binding protein TF1 in solution, *J. Mol. Biol.* 263, 259-68.
- Jones, B. E., Dossonnet, V., Kuester, E., Hillen, W., Deutscher, J., and Klevit, R. E. (1997). Binding of the catbolite repressor protein CcpA to its DNA target is regulated by phosphorylation of its corepressor HPr, *J. Biol. Chem.* 272, 26530-26535.
- Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999). Protein-DNA interactions: A structural analysis, *J. Mol. Biol.* 287, 877-96.
- Jones, T. A. (1985). Interactive computer graphics: FRODO, *Methods Enzymol.* 115, 157-171.
- Jones, T. A. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr.* A47, 110-119.

- Jordan, S. R., and Pabo, C. O. (1988). Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions, *Science* 242, 893-899.
- Kaptein, R., Zuderweg, E. R. P., Scheek, R. M., Boelens, R., and van Gunsteren, W. (1985). A protein structure from nuclear magnetic resonance data: *lac* repressor headpiece, *J. Mol. Biol.* 182, 179.
- Kato, M., Aiba, H., Tate, S., Nishimura, Y., and Mizuno, T. (1989). Location of phosphorylation site and DNA-binding site of a positive regulator, OmpR, involved in activation of the osmoregulatory genes of *Escherichia coli*, *FEBS Lett.* 249, 168-72.
- Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (1960). Structure of myoglobin, *Nature* 185, 422-427.
- Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993a). Co-crystal structure of TBP recognizing the minor groove of a TATA element, *Nature* 365, 520-7.
- Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993b). Crystal structure of a yeast TBP/TATA-box complex, *Nature* 365, 512-20.
- Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search, *Acta Crystallogr. D Biol. Crystallogr.* D55, 484-491.
- Kobe, B., and Kemp, B. E. (1999). Active site-directed protein regulation, *Nature* 402, 373-6.
- Kostrewa, D., Granzin, J., Koch, C., Choe, H.-W., Raghunathan, S., and Wolf, W. (1991). Three-dimensional structure of the *E. coli* DNA-binding protein, FIS, *Nature* 349, 178.
- Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection, *Proc. Natl. Acad. Sci. U.S.A.* 82, 488-92.

- Kwon, H. J., Bennik, M. H., Demple, B., and Ellenberger, T. (2000). Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA, *Nat. Struct. Biol.* 7, 424-30.
- Lanio, T., Jeltsch, A., and Pingoud, A. (2000). On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: a case study employing *EcoRV* as the target, *Protein Eng.* 13, 275-281.
- Lanzilotta, W. N., Schuller, D. J., Thorsteinsson, M. V., Kerby, R. L., Roberts, G. P., and Poulos, T. L. (2000). Structure of the CO sensing transcription activator CooA, *Nat. Struct. Biol.* 7, 876-80.
- Laskowski, R. A., MacArthur, M. W., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallog.* 26, 283-291.
- Lavery, R., and Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids, *J. Biomol. Struct. Dyn.* 6, 63-91.
- Lawson, C. L., Zhang, R. G., Schevitz, R. W., Otwinowski, Z., Joachimiak, A., and Sigler, P. B. (1988). Flexibility of the DNA-binding domains of trp repressor, *Proteins* 3, 18-31.
- Lee, J. H., and Hoover, T. R. (1995). Protein crosslinking studies suggest that *Rhizobium meliloti* C4-dicarboxylic acid transport protein D, a  $\sigma^{54}$ -dependent transcriptional activator, interacts with  $\sigma^{54}$  and the  $\beta$  subunit of RNA polymerase, *Proc. Natl. Acad. Sci. U.S.A.* 92, 9702-9706.

- Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer, *Science* 271, 1247-54.
- Lewis, R. J., Brannigan, J. A., Offen, W. A., Smith, I., and Wilkinson, A. J. (1998). An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex, *J. Mol. Biol.* 283, 907-12.
- Li, M., Moyle, H., and Susskind, M. (1994). Target of the transcriptional activation function of phage  $\lambda$  cI protein, *Science* 263, 75-77.
- Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). *Science* 270, 262-269.
- Liu, K., and Hanna, M. M. (1995). NusA interferes with interactions between the nascent RNA and the C-terminal domain of the alpha subunit of RNA polymerase in *Escherichia coli* transcription complexes, *Proc. Natl. Acad. Sci. U.S.A.* 92, 5012-5016.
- Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites, *J. Mol. Biol.* 277, 2177-2198.
- Lu, F. (1999), Ph.D. Dissertation, Purdue University, Purdue, IN.
- Lu, F., Brennan, R. G., and Zalkin, H. (1998a). *Escherichia coli* purine repressor: key residues for the allosteric transition between active and inactive conformations and for interdomain signaling, *Biochemistry* 37, 15680-90.
- Lu, F., Schumacher, M. A., Arvidson, D. N., Haldimann, A., Wanner, B. L., Zalkin, H., and Brennan, R. G. (1998b). Structure-based redesign of corepressor specificity of the *Escherichia coli* purine repressor by substitution of residue 190, *Biochemistry* 37, 971-82.

- Luisi, B. F. (1995). DNA-protein interaction at high resolution. In DNA-protein: structural interactions, D. M. J. Lilley, ed. (Oxford, IRL Press), pp. 1-48.
- Lundblad, J. R., Laurance, M., and Goodman, R. H. (1996). Fluorescence polarization analysis of protein-DNA and protein-protein interactions, *Mol Endocrinol* 10, 607-12.
- Mandel-Gutfreund, Y., Schueler, O., and Margalit, H. (1995). Comprehensive analysis of hydrogen bonds in regulatory protein-DNA complexes: in search of common principles, *J. Mol. Biol.* 253, 370-388.
- Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., and Miller, J. H. (1994). Genetic studies of the lac repressor. XIV. Analysis of 4000 altered *Escherichia coli* lac repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence, *J. Mol. Biol.* 240, 421-33.
- Martinez-Hackert, E., and Stock, A. M. (1997). The DNA-binding domain of OmpR: crystal structures of a winged helix transcription factor, *Structure* 5, 109-24.
- Matthews, B. W. (1996). Structural and genetic analysis of the folding and function of T4 lysozyme, *FASEB J* 10, 35-41.
- Matthews, K. S., and Nichols, J. C. (1998). Lactose repressor protein: functional properties and structure, *Prog. Nucl. Acid Res. Mol. Biol.* 58, 127-164.
- Maxam, A., and Gilbert, W. (1977). A new method for sequencing DNA, *Proc. Natl. Acad. Sci. U.S.A.* 74, 560-564.
- McKay, D. B., and Steitz, T. A. (1981). Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA, *Nature* 290, 744-9.

- McKay, D. B., Weber, I. T., and Steitz, T. A. (1982). Structure of catabolite gene activator protein at 2.9 Å resolution: incorporation of amino acid sequence and interactions with cAMP, *J. Biol. Chem.* 257, 9518.
- McPherson, A. (1999). Crystallization of biological macromolecules, 1st edn (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).
- McRee, D. E. (1992). A visual protein crystallographic software system for X11/XView, *J. Mol. Graphics* 10, 44-46.
- Meibom, K. L., Kallipolitis, B. H., Ebright, R. H., and Valentin-Hansen, P. (2000). Identification of the subunit of cAMP receptor protein (CRP) that functionally interacts with CytR in CRP-CytR-mediated transcriptional repression, *J. Biol. Chem.* 275, 11951-6.
- Meng, L. M., Kilstrup, M., and Nygaard, P. (1990). Autoregulation of PurR repressor synthesis and involvement of *purR* in the regulation of *purB*, *purC*, *purL*, *purMN*, and *guaBA* expression in *Escherichia coli*, *Eur. J. Biochem.* 187, 373-379.
- Miller, A., Wood, D., Ebright, R. H., and Rothman-Denes, L. B. (1997). RNA polymerase  $\beta'$  subunit: a target of DNA-binding-independent activation, *Science* 275, 1655-1657.
- Miller, J. H. (1972). Experiments in Molecular Genetics (Cold Spring Harbor, Cold Spring Harbor Laboratory).
- Mills, S. D., Jasalavich, C. A., and Cooksey, D. O. (1993). A two-component regulatory system required for copper-inducible expression of the copper resistance operon of *Pseudomonas syringae*, *J. Bacteriol.* 175, 1656-64.



- Mironov, A. S., Nechayeva, G. D., and Sukhodolets, V. V. (1989). Interaction of negative (CytR) and positive (cAMP-CRP) control of the promoter region of the uridine phosphorylase (*udp*) gene in *Escherichia coli* K-12, *Genetica* 25, 438-447.
- Misra, T. K. (1992). Bacterial resistance to inorganic mercury salts and organomercurials, *Plasmid* 27, 4-16.
- Mobley, H. L., Island, M. D., and Massad, G. (1994). Virulence determinants of uropathogenic *Escherichia coli* and *Proteus mirabilis*, *Kidney Int. Suppl.* 47, S129-36.
- Mondragon, A., Subbiah, S., Almo, S. C., Drottar, M., and Harrison, S. C. (1989a). Structure of the amino-terminal domain of phage 434 repressor at 2.0 Å resolution, *J. Mol. Biol.* 205, 189-200.
- Mondragon, A., Wolberger, C., and Harrison, S. C. (1989b). Structure of phage 434 Cro protein at 2.35 Å resolution, *J. Mol. Biol.* 205, 179-88.
- Monod, J., Wyman, J., and Changeaux, J.-P. (1965). On the nature of allosteric transitions: a plausible model, *J. Mol. Biol.* 12, 88.
- Mowbray, S. L., and Cole, L. B. (1992). 1.7 Å X-ray structure of the periplasmic ribose receptor from *E. coli*, *J. Mol. Biol.* 225, 155-175.
- Munagala, N., Sarver, A. E., and Wang, C. C. (2000). Converting the guanine phosphoribosyltransferase from *Giardia lamblia* to a hypoxanthine-guanine phosphoribosyltransferase, *J. Biol. Chem.* 275, 37072-37077.
- Munson, G. P., Lam, D. L., Outten, F. W., and O'Halloran, T. V. (2000). Identification of a copper-responsive two-component system on the chromosome of *Escherichia coli* K-12, *J. Bacteriol.* 182, 5864-71.

- Nadassy, K., Wodak, S. J., and Janin, J. (1999). Structural features of protein-nucleic acid recognition sites, *Biochemistry* 38, 1999-2017.
- Nagadoi, A., Morikawa, S., Nakamura, H., Enari, M., Kobayashi, K., Yamamoto, H., Sampei, G., Mizobuchi, K., Schumacher, M. A., Brennan, R. G., and et al. (1995). Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain, *Structure* 3, 1217-24.
- Nelson, H. C. M., Finch, J. T., Luisi, B. F., and Klug, A. (1987). The structure of an oligo(dA)•oligo(dT) tract and its biological implications, *Nature* 330, 221-226.
- Nguyen, C. C., and Saier, J., M.H. (1995). Phylogenetic, structural, and functional analyses of the LacI-GalR family of bacterial transcription factors, *FEBS Lett.* 377, 98-102.
- Nies, D. H. (1992). Resistance to cadmium, cobalt, zinc, and nickel in microbes, *Plasmid* 27, 17-28.
- Nies, D. H., and Silver, S. (1989). Plasmid-determined inducible efflux is responsible for resistance to cadmium, zinc, and cobalt in *Alcaligenes eutrophus*, *J. Bacteriol.* 171, 896-900.
- Nilsson, L. O., Gustafsson, A., and Mannervik, B. (2000). Redesign of substrate-selectivity determining modules of glutathione transferase A1-1 installs high catalytic efficiency with toxic alkenyl products of lipid peroxidation, *Proc. Natl. Acad. Sci. U.S.A.* 97, 9408-12.
- Niu, W., Zhou, Y., Dong, Q., Ebright, Y. W., and Ebright, R. H. (1994). Characterization of the activating region of *Escherichia coli* catabolite gene activator protein (CAP) I. Saturation and alanine scanning mutagenesis, *J. Mol. Biol.* 243, 595-602.

- Noll, M., Petrukhin, K., and Lutsenko, S. (1998). Identification of a novel transcription regulator from *Proteus mirabilis*, PMTR, revealed a possible role of YJAI protein in balancing zinc in *Escherichia coli*, *J. Biol. Chem.* 273, 21393-401.
- O'Halloran, T. V. (1993). Transition metals in control of gene expression, *Science* 261, 715-25.
- O'Halloran, T. V., Frantz, B., Shin, M. K., Ralston, D. M., and Wright, J. G. (1989). The MerR heavy metal receptor mediates positive activation in a topologically novel transcription complex, *Cell* 56, 119-29.
- Odermatt, A., and Solioz, M. (1995). Two trans-acting metalloregulatory proteins controlling expression of the copper-ATPases of *Enterococcus hirae*, *J. Biol. Chem.* 270, 4349-54.
- Okamura, H., Hanaoka, S., Nagadoi, A., Makino, K., and Nishimura, Y. (2000). Structural comparison of the PhoB and OmpR DNA-binding/transactivation domains and the arrangement of PhoB molecules on the phosphate box, *J. Mol. Biol.* 295, 1225-36.
- Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system, *Nat. Struct. Biol.* 7, 215-9.
- Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution, *Nature* 335, 321-9.
- Outten, C. E., Outten, F. W., and O'Halloran, T. V. (1999). DNA distortion mechanism for transcriptional activation by ZntR, a Zn(II)-responsive MerR homologue in *Escherichia coli*, *J. Biol. Chem.* 274, 37517-24.

- Outten, F. W., Outten, C. E., Hale, J., and O'Halloran, T. V. (2000). Transcriptional activation of an *Escherichia coli* copper efflux regulon by the chromosomal MerR homologue, CueR, J. Biol. Chem. 275, 31024-9.
- Pabo, C. O., and Lewis, M. (1982). The operator-binding domain of  $\lambda$  repressor: structure and DNA recognition, Nature 298, 443-447.
- Parkhill, J., and Brown, N. L. (1990). Site-specific insertion and deletion mutants in the *mer* promoter-operator region of Tn501; the nineteen base-pair spacer is essential for normal induction of the promoter by MerR, Nucleic Acids Res. 18, 5157-5162.
- Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. H., and Berman, H. M. (1996). Structure of the CAP-DNA complex at 2.5 Å resolution: a complete picture of the protein-DNA interface, J. Mol. Biol. 260, 395-408.
- Peet, D. J., Doyle, D. F., Corey, D. R., and Mangelsdorf, D. J. (1998). Engineering novel specificities for ligand-activated transcription in the nuclear hormone receptor RXR, Chem. Biol. 5, 13-21.
- Pelton, J. G., Kustu, S., and Wemmer, D. E. (1999). Solution structure of the DNA-binding domain of NtrC with three alanine substitutions, J. Mol. Biol. 292, 1095-1110.
- Penin, F., Geourjon, C., Montserret, R., Bockmann, A., Lesage, A., Yang, Y. S., Bonod-Bidaud, C., Cortay, J. C., Negre, D., Cozzzone, A. J., and Deleage, G. (1997). Three-dimensional structure of the DNA-binding domain of the fructose repressor from *Escherichia coli* by 1H and 15N NMR, J. Mol. Biol. 270, 496-510.
- Pohl, E., Holmes, R. K., and Hol, W. G. (1999). Crystal structure of the iron-dependent regulator (IdeR) from *Mycobacterium tuberculosis* shows both metal binding sites fully occupied, J. Mol. Biol. 285, 1145-56.

- Quioco, F. A., and Ledvina, P. S. (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes, *Mol. Microbiol.* 20, 17-25.
- Quioco, F. A., and Vyas, N. K. (1984). Novel stereospecificity of the L-arabinose-binding protein, *Nature* 310, 381-386.
- Rafferty, J. B., Somers, W. S., Saint-Girons, I., and Phillips, S. E. (1989). Three-dimensional crystal structures of *Escherichia coli met* repressor with and without corepressor, *Nature* 341, 705-10.
- Ralston, D. M., and O'Halloran, T. V. (1990). Metalloregulatory proteins and molecular mechanisms of heavy metal signal transduction, *Adv. Inorg. Biochem.* 8, 1-31.
- Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding, *Nature* 362, 219-23.
- Rashin, A. A., Rashin, B. H., Rashin, A., and Abagyan, R. (1997). Evaluating the energetics of empty cavities and internal mutations in proteins, *Protein Sci.* 6, 2143-2158.
- Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994). DNA recognition by b-sheets in the Arc repressor-operator crystal structure, *Nature* 367, 754-757.
- Rensing, C., Mitra, B., and Rosen, B. P. (1998). A Zn(II)-translocating P-type ATPase from *Proteus mirabilis*, *Biochem Cell Biol* 76, 787-90.
- Rhee, S., Martin, R. G., Rosner, J. L., and Davies, D. R. (1998). A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator, *Proc. Natl. Acad. Sci. U.S.A.* 95, 10413-8.

- Rice, P. A., Yang, S., Mizuuchi, K., and Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn, *Cell* 87, 1295-306.
- Richardson, J. S., and Richardson, D. C. (1988). Amino acid preference for specific locations at the ends of helices, *Science* 240, 1648-52.
- Rolfes, R. J., and Zalkin, H. (1988a). *Escherichia coli* gene *purR* encoding a repressor protein for purine nucleotide synthesis. Cloning, nucleotide sequence, and interaction with the *purF* operator, *J. Biol. Chem.* 263, 19653-61.
- Rolfes, R. J., and Zalkin, H. (1988b). Regulation of *Escherichia coli purF*. Mutations that define the promoter, operator, and purine repressor gene, *J. Biol. Chem.* 263, 19649-52.
- Rolfes, R. J., and Zalkin, H. (1990a). Autoregulation of *Escherichia coli purR* requires two control sites downstream of the promoter, *J. Bacteriol.* 172, 5758-66.
- Rolfes, R. J., and Zalkin, H. (1990b). Purification of the *Escherichia coli* purine regulon repressor and identification of corepressors, *J. Bacteriol.* 172, 5637-42.
- Rosen, B. P. (1999). The role of efflux in bacterial resistance to soft metals and metalloids, *Essays Biochem.* 34, 1-15.
- Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase, *Science* 262, 1407-13.
- Rutherford, J. C., Cavet, J. S., and Robinson, N. J. (1999). Cobalt-dependent transcriptional switching by a dual-effector MerR-like protein regulates a cobalt-exporting variant CPx-type ATPase, *J. Biol. Chem.* 274, 25827-25832.

- Sack, J. S., Saper, M. A., and Quioco, F. A. (1989). Periplasmic binding protein structure and function: refined structures of the leucine/isoleucine/valine-binding protein and its complex with leucine, *J. Mol. Biol.* *206*, 171-191.
- Sarai, A., and Takeda, Y. (1989).  $\lambda$  repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically, *Proc. Natl. Acad. Sci. U.S.A.* *86*, 6513-17.
- Savery, N. J., Lloyd, G. S., Kainz, M., Gaal, T., Ross, W., Ebright, R. H., Gourse, R. L., and Busby, S. J. (1998). Transcription activation at Class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase  $\alpha$  subunit, *EMBO J.* *17*, 3439-47.
- Schiering, N., Tao, X., Zeng, H., Murphy, J. R., Petsko, G. A., and Ringe, D. (1995). Structures of the apo- and the metal ion-activated forms of the diphtheria *tox* repressor from *Corynebacterium diphtheriae*, *Proc. Natl. Acad. Sci. U.S.A.* *92*, 9843-50.
- Schildbach, J. F., Karzai, A. W., Raumann, B. E., and Sauer, R. T. (1999). Origins of DNA-binding specificity: role of protein contacts with the DNA backbone, *Proc. Natl. Acad. Sci. U.S.A.* *96*, 811-7.
- Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees, *Science* *253*, 1001-1007.
- Schumacher, M. A., Choi, K. Y., Lu, F., Zalkin, H., and Brennan, R. G. (1995). Mechanism of corepressor-mediated specific DNA binding by the purine repressor, *Cell* *83*, 147-55.

- Schumacher, M. A., Choi, K. Y., Zalkin, H., and Brennan, R. G. (1992). Crystallization and preliminary X-ray studies on the co-repressor binding domain of the *Escherichia coli* purine repressor, J. Mol. Biol. 225, 1131-3.
- Schumacher, M. A., Choi, K. Y., Zalkin, H., and Brennan, R. G. (1994a). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by  $\alpha$  helices, Science 266, 763-70.
- Schumacher, M. A., Choi, K. Y., Zalkin, H., and Brennan, R. G. (1994b). Crystallization and preliminary X-ray analysis of an *Escherichia coli* purine repressor-hypoxanthine-DNA complex, J. Mol. Biol. 242, 302-5.
- Schumacher, M. A., Glasfeld, A., Zalkin, H., and Brennan, R. G. (1997). The X-ray structure of the PurR-guanine-*purF* operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity, J. Biol. Chem. 272, 22648-53.
- Schumacher, M. A., Macdonald, J. R., Bjorkman, J., Mowbray, S. L., and Brennan, R. G. (1993). Structural analysis of the purine repressor, an *Escherichia coli* DNA-binding protein, J. Biol. Chem. 268, 12282-8.
- Seeman, N. C., Roseberg, J. M., and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins, Proc. Natl. Acad. Sci. U.S.A. 73, 804-808.
- Sevilla-Sierra, P., Otting, G., and Wuthrich, K. (1994). Determination of the nuclear magnetic resonance structure of the DNA-binding domain of the P22 c2 repressor (1 to 76) in solution and comparison with the DNA-binding domain of the 434 repressor, J. Mol. Biol. 235, 1003-20.
- Silver, S. (1996). Bacterial resistances to toxic metal ions--a review, Gene 179, 9-19.



- Utschig, L. M., Bryson, J. W., and O'Halloran, T. V. (1995). Mercury-199 NMR of the metal receptor site in MerR and its protein-DNA complex, *Science* 268, 380-5.
- Valentin-Hansen, P., Sogaard-Andersen, L., and Pedersen, H. (1996). A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control, *Mol. Microbiol.* 20, 461-6.
- Vlassi, M., Cesareni, G., and Kokkinidis, M. (1999). A correlation between the loss of hydrophobic core packing interactions and protein stability, *J. Mol. Biol.* 285, 817-827.
- Vos, S., de Jersey, J., and Martin, J. L. (1997). Crystal structure of *Escherichia coli* xanthine phosphoribosyltransferase, *Biochemistry* 36, 4125-34.
- Wang, Y., Zhao, S., Somerville, R. L., and Jardetzky, O. (2001). Solution structure of the DNA-binding domain of the TyrR protein of *Haemophilus influenzae*, *Protein Sci.* 10, 592-598.
- White, A., Ding, X., vanderSpek, J. C., Murphy, J. R., and Ringe, D. (1998). Structure of the metal-ion-activated diphtheria toxin repressor/*tox* operator complex, *Nature* 394, 502-6.
- White, S. W., Wilson, K. S., Appelt, K., and Tanaka, I. (1999). The high-resolution structure of DNA-binding protein HU from *Bacillus stearothermophilus*, *Acta Crystallogr. D Biol. Crystallogr.* D55, 801-9.
- Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J., and Matthews, B. W. (1992). *Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains, *Proc. Natl. Acad. Sci. U.S.A.* 89, 9257-61.

- Wintjens, R., and Rooman, M. (1996). Structural classification of HTH DNA-binding domains and protein-DNA interaction modes, *J. Mol. Biol.* 262, 294-313.
- Wojciak, J. M., Iwahara, J., and Clubb, R. T. (2001). The Mu repressor-DNA complex contains an immobilized 'wing' within the minor groove, *Nat. Struct. Biol.* 8, 84-90.
- Wolberger, C., Dong, Y. C., Ptashne, M., and Harrison, S. C. (1988). Structure of a phage 434 Cro/DNA complex, *Nature* 335, 789.
- Wouters, J. (1998). Cation- $\pi$  ( $\text{Na}^+$ -Trp interactions in the crystal structure of tetragonal lysozyme, *Protein Sci.* 7, 2472-2475.
- Wu, W. F., Urbanowski, M. L., and Stauffer, G. V. (1993). MetJ-mediated regulation of the *Salmonella typhimurium metE* and *metR* genes occurs through a common operator region, *FEMS Microbiol. Lett.* 108, 145-50.
- Yang, C.-C., and Nash, H. A. (1989). The interaction of *E. coli* IHF protein with its specific binding sites, *Cell* 57, 869-880.
- Zalkin, H., and Nygaard, P. (1996). Biosynthesis of purine nucleotides. In *Escherichia coli* and *Salmonella*, cellular and molecular biology, R. C. I. F. C. Neidhardt, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, eds.
- Zhou, Y., Busby, S., and Ebright, R. H. (1993). Identification of the functional subunit of a dimeric transcription activator protein by use of oriented heterodimers, *Cell* 73, 375-9.