

DNA RECOGNITION BY PROKARYOTIC TRANSCRIPTIONAL REGULATORS

by

Katherine Metcalf\_Hoffmann

A DISSERTATION

Presented to the Department of Biochemistry and Molecular Biology

and the Oregon Health & Science University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

August 2006

School of Medicine  
Oregon Health Sciences University

---

Certificate of Approval

---

This is to certify that the Ph.D. thesis of  
Katherine Metcalf Hoffmann  
has been approved

[Redacted Signature]

---

Professor in charge of thesis

[Redacted Signature]

---

Member

[Redacted Signature]

---

Member

[Redacted Signature]

---

Member

[Redacted Signature]

---

Member

<b>TABLE OF CONTENTS</b>	<b>page #</b>
<b>List of tables and figures</b>	vi
<b>Acknowledgements</b>	viii
<b>Abstract</b>	1
<b>Chapter 1. Introduction</b>	
Transcriptional regulation	3
Overview of Protein/DNA interactions	
Direct vs. Indirect Readout	4
The first structure of a transcriptional regulator: TrpR	5
Overview of indirect readout in notable protein/DNA complexes	6
CAP/DNA	6
TBP/DNA	6
434 Repressor	7
MetJ	8
The Purine Repressor (PurR) in <i>E. coli</i> .	
Master transcriptional regulation of <i>de novo</i> purine synthesis, also purine salvage and pyrimidine <i>de novo</i> biosynthesis.	9
LacI family function	10
PurR structures and conformational changes	11
LacI family structures	
DNA structure and deformability	13
The Multiple Transferable Resistance Regulator (MtrR) from <i>N. gonorrhoeae</i>	
TetR family	15
TetR family structures	
TetR bound and unbound to DNA	16
QacR apo, drug bound and DNA bound structures	18
EthR structural information	22
MtrR discovery and function	
Circuit, binding site theories	23
Homology to other TetR family members	24
Tables and Figures	28
<b>Chapter 2. Methods</b>	
Macromolecular crystallography	46
Crystal growth	
Solubility manipulation	46
Nuclei manipulation (seeding)	47
Crystal symmetry	48
Crystal systems	48
Bravais lattices	49
Space groups (statistical disorder)	50
X-ray diffraction	51

Coherent scattering	51
Bragg's Law	52
Reciprocal space	52
The structure factor (Fourier series)	53
The phase problem	54
Molecular modification	54
<i>de novo</i> phase solution methods	55
Multiple isomorphous replacement	55
Multiple anomalous dispersion	56
Molecular replacement	56
Refinement	58
Rigid body and resolution cutoffs	58
xyz and B-factor refinement	59
R-factors as a measure of a model's fitness to data	60
Composite omit maps	60
Ramachandran plots	61
Fluorescence Spectroscopy	62
Fluorescence Polarization	62
Circular Dichroism	63
Dynamic Light Scattering	64
Tables and Figures	66
<b>Chapter 3. Transcriptional Regulation and Specificity for Cognate DNA Through Indirect Readout Demonstrated in the Purine Repressor (PurR) Complex</b>	<b>77</b>
Abstract	78
Introduction	79
Materials and methods	84
Results	87
Discussion	94
Tables and Figures	99
<b>Chapter 4. Characterization of the Multiple Transferable Resistance Repressor, MtrR, from <i>Neisseria gonorrhoeae</i></b>	<b>107</b>
Abstract	108
Introduction	109
Methods and Results	111
Tables and Figures	118
<b>Chapter 5. The Role of a Basic N-Terminal Extension of the <i>Neisseria gonorrhoeae</i> Multiple Transferable Resistance Repressor, MtrR, in DNA Binding Affinity and Specificity.</b>	<b>123</b>
Abstract	124
Introduction	125
Materials and methods	129
Results	132
Tables and Figures	141

**Chapter 6. Summary and Conclusions**

**152**

**Literature Cited.**

**155**

## List of Tables and Figures:

### Chapter 1.

Figure 1.01: Ribbon model of PurR/hypoxanthine/*purF* complex.

Figure 1.02: The Tryptophan Repressor (TrpR)-operator complex

Figure 1.03: The Catabolite Activator Protein (CAP)-DNA complex.

Figure 1.04: The TATA-binding protein (TBP) -operator complex.

Figure 1.05: Two oligomers of the methionine repressor (MetJ) bound to two operator metboxes.

Figure 1.06: The *pur* operon. A) Alignment and homology among the full-site PurR binding sites. B) genetic structure of the promoters and effective repression by PurR.

Figure 1.07: The LacI DNA binding domain complexed with DNA.

Figure 1.08: The PurR ligand binding pocket.

Figure 1.09: Cartoon demonstration of A-tracts as rigid rods in mixed sequences.

Figure 1.10: PurR-DNA contacts in a single *purF* half-site.

Figure 1.11: Cartoon representation of purine and pyrimidine base pairs in the *purF* halfsite demonstrating intra-base pair propellor twist, and the inter-base pair twist and roll.

Figure 1.12: Alignment of selected TetR family member's DNA binding domains.

Figure 1.13: Tetracycline and the Tetracycline repressor in apo and DNA bound structures.

Figure 1.14: QacR structures, bound to DNA and bound to two drugs in the multidrug binding pocket.

Figure 1.15: The apo crystal structures of two other TetR family members, EthR and CprB.

Figure 1.16: A schematic representation of *mtrR* and *mtrC* detailing the proposed MtrR binding sites of the AT rich inverted repeat (IR) and the 31 base pair imperfect direct repeat (DR) sequence pinpointed in footprinting assays.

Figure 1.17: A model of the 8 residue N-terminal extension of MtrR on the QacR DNA bound dimer of dimers.

## Chapter 2.

Figure 2.01: A saturation diagram for macromolecule crystal growth.

Figure 2.02 Definitions of a unit cell and the parameters and symmetry that defines cell types.

Figure 2.03: The fourteen Bravais lattices.

Figure 2.04: An illustration of Bragg's Law.

Figure 2.05: Constructing reciprocal space from real space.

Figure 2.06. The Ewald sphere.

Figure 2.07. Phase determination by the Harker construction.

Figure 2.08. Ramachandran plots.

Figure 2.09. The chemical structure of fluorescein, and the spin associated with binding in an anisotropy experiment.

Figure 2.10. The interaction of a circularly polarized wave with a helical molecule.

Figure 2.11. Circular Dichroism spectra for protein secondary structure and DNA.

### **Chapter 3.**

Table 3.01: Selected Crystallographic Data and Statistics.

Table 3.02: DNA Structural Analysis and Selected Structural and Functional Statistics.

Table 3.03: Analysis of Protein Contacts (Hydrogen Bond Distances).

Figure 3.01: Comparative Binding Isotherms for PurR binding to the single and double half-site variants of *purF*.

Figure 3.02: A) Overlay of full-length PurR/hypoxanthine/*purF* structure and demonstrated differences in global DNA structures.

Figure 3.03: Closeup on the substituted DNA bases overlaid on the wild-type Adenine.

### **Chapter 4**

Table 4.01. Secondary structure content (%) of MtrR and QacR.

Figure 4.01. Salt effect and site length binding isotherms of MtrR and QacR.

Figure 4.02. Determination of the stoichiometry of MtrR-DNA binding.

Figure 4.03. Circular dichroism spectra of MtrR and QacR.

### **Chapter 5**

Table 5.01. Binding constants for all MtrR mutants and all DNA sites reported in this paper.

Figure 5.01. Aligned sequences of the N-terminal extensions of 30 TetR sub-family members.

Figure 5.02. Wild-type MtrR Binding to DNA Sites Proposed in the Literature.

Figure 5.03. Salt Effect in MtrR Deletion Mutants and Specificity for DNA Sequence.

Figure 5.04. Binding Affinity of MtrR Site Mutants To the Full Length Direct Repeat  
DNA Site.

Fig. 5.05. The eight residue N-terminal extension of MtrR was modeled onto helix a1 of  
QacR bound to DNA (1QPZ).

## **Acknowledgements:**

First, I am indebted to Dr. Brennan for, among other things, being not only the first person I interacted with while considering whether to come to graduate school at OHSU, but also being with me as I finish my dissertation; he has been an wonderful mentor. I am also grateful to Dr. Svetlana Lutsenko and the other members of my committee for their patience and guidance.

In the course of graduate school, several other extraordinary people spent a great deal of time and energy answering my questions. One was Dr. Joy Huffman, who was and is an amazing listener and teacher; Drs. Greg Allen, Kate Newberry and Maria Schumacher also were generous with their time and knowledge. I am particularly indebted to Maria for laying the groundwork for the projects I worked on, and for remembering the details of them so meticulously that she could call my attention to overlooked areas from memory. I'd like to thank Dr. William Shafer and Dr. Hans Peter Bächinger for their generosity as collaborators in lending their time, their people, their supplies and their equipment, and for being so prompt responding to requests; specifically, I worked with Jaqueline Pickel and Eric Steele, and I am very grateful for their time, energy and expertise. Jaqueline provided the  $\Delta 8$  and  $\Delta 10$  MtrR mutants, as well as several wild-type constructs for that project; In the Brennan lab, Robert Pearson provided the S53A PurR protein used in the crystallography studies, and I am grateful to both of them for that very material assistance.

I'd like to dedicate this thesis to my family, who after many years of rolling their eyes when I talked about how excited I was about science, have finally figured out how to describe my work to their friends, only to have me move on to new projects. I'm so pleased that everyone is happy and healthy and here to see it.

I'd also like to dedicate this thesis to Dr. David Murray (1978-2005), who didn't get to participate in this part.

## **Abstract**

Regulation of protein production in a cell is critical both for the efficient management of resources and for maintaining balance of productive proteins that, if over expressed, could cause damage. Enzymes involved in purine and pyrimidine biosynthesis, for example, would be energetically wasteful in the absence of the building blocks needed to create the nucleosides. Alternatively, a multidrug export pump that utilizes a proton gradient to extrude potentially hazardous antibacterial agents could deplete the gradient it makes use of by exporting non-hazardous agents undiscerning if it were present in greater levels than necessary.

For this reason, the transcription of DNA, the levels of mRNA, post-transcriptional modifications, translation, post-translational modification and protein degradation are all regulated within a cell to maintain the appropriate levels of particular proteins in response to tailored stimuli; of these, transcriptional regulation seems the tightest regulated, likely because the most efficient way to regulate levels of protein is to control how much is created from the very beginning.

Transcriptional regulatory proteins may bind DNA for the purpose of either repressing or activating transcription of a target gene. The purpose of this thesis is to explore the nature of the interaction between DNA binding proteins and DNA using as model systems two bacterial transcriptional regulators: the Multiple Transferable Resistance Regulator (MtrR) from *Neisseria gonorrhoeae* and the Purine Repressor (PurR) from *Escherichia coli* using a variety of biological, biochemical and biophysical techniques as

appropriate. The experimental methods used in these experiments and the science behind them are detailed in chapter two allowing for minimal discussion of methods in the experimental chapters. Of the three structure-function papers included as chapters in the thesis, the first presents PurR in a structural exploration of direct and indirect readout and the last two describe MtrR; primary characterization and further detailed exploration of the nature of the DNA binding domain function. The final chapter will conclude the observations gleaned from these papers taken as a whole.

## Chapter 1: Introduction

### Overview of protein/DNA interactions

Prokaryotic promoters consist of the upstream -10 (TATAA) and -35 (TTGACA) boxes and occasionally (increasing the efficiency of certain promoters) an additional AT-rich ~20 bp upstream (UP) element, bound by the  $\alpha$  protein of the core RNA polymerase (RNAP) which also comprises an additional  $\alpha$ ,  $\beta$ ,  $\beta'$ , and finally  $\sigma$ , completing the holoenzyme and allowing it to recognize the -35 and -10 boxes as the polymerase slides along the promoter.

This binding of RNAP to the promoter element of a gene is the first stage of initiation of transcription while DNA is double stranded and the protein is in its 'closed' form. An isomerization of RNAP proteins allows the local melting of DNA to single stranded template and the 'open' form, the first phosphodiester bonds of the RNA transcript are formed, and finally the entire complex clears the promoter entirely and transitions to the gene proper, allowing  $\sigma$  to dissociate and elongation of the transcript until termination. Each step of transcription can be regulated, but to conserve maximal energy in a cell, initiation is the most highly regulated.

Transcriptional regulation may come in the form of either activation or repression of transcription when the regulator is present, and is a result of any of a number of mechanisms to affect a result. An activator might upregulate the transcription of its target gene by distorting or bending the DNA into a more favorable form, allowing

RNAP to bind, as is the case with several members of the MerR family of transcriptional regulators. Specifically, BmrR (*B. subtilis* multidrug resistance regulator, a MerR family member) binds an unusually long 19 base pair inverted repeat in the region between its target gene's -10 and -35 boxes, overtwisting the DNA in the process such that the two boxes are reoriented from opposite faces of the DNA to the same face; a much more favorable conformation for RNAP to bind. Conversely, a repressor protein might physically interfere with RNAP initiation by blocking RNAPs access to the promotor, as PurR does by binding the -10 box first; alternatively, a repressor may prevent RNAPs transition to the translating form, as QacR (quaternary ammonium compound regulator) seems to, binding as it does just downstream from the -10 box of its target gene.

Inherent in this ability to recruit, block or inhibit RNAPs association with the DNA, however, is the ability to very specifically bind and manipulate particular pieces of DNA in response to specific environmental signal. PurR, for example, does not bind DNA without first binding corepressor (figure 1.01), an end-product signal from the cell to stop further purine biosynthesis. This signal triggers conformational changes in the protein allowing for both direct readout of the sequence of DNA base pairs at its binding site, including amino acid-to-base electrostatic interactions, hydrogen bonding, Van Der Waals contacts and exclusions, in contacts to both the major and the minor grooves. Additionally, there is indirect readout of the bases in the binding site, as the protein makes backbone and phosphate contacts, reads the deformability of the DNA and kinks and bends it in specific places. These techniques as well as other methods of DNA interaction including DNA warping, twisting or unwinding are used by transcriptional

regulators to effect specific interactions with specific target DNA. Despite clear evidence that both direct and indirect readout of DNA contributes to the specificity of DNA recognition, it has been easier to visualize direct readout in structural studies.

The first transcriptional regulator to be crystallized and solved in complex with its cognate DNA to high resolution was the Tryptophan Repressor (TrpR) in 1988 by Otwinowski *et al.* The TrpR-operator complex was solved to 2.4 Å resolution and revealed an extensive contact surface including 24 direct van der Waals contacts between the  $\alpha$ -helical protein and DNA, and 6 solvent mediated hydrogen bonds to the phosphates of the DNA backbone (figure 1.02). There were, surprisingly, no direct hydrogen bonds or non-polar contacts to the bases that explained the specificity of the protein for the operator. (Zhang, Joachimiak *et al.* 1987; Otwinowski, Schevitz *et al.* 1988)

With the expected chemical component of specificity (direct readout) making significantly less of a contribution to binding, it was concluded that the steric component (indirect readout) was the greater contributor to the affinity and specificity of TrpR for its operator, where the geometry of the phosphate backbone permits a stable interface and the water mediated polar contacts are the only contacts to the bases themselves (Zhang, Joachimiak *et al.* 1987; Otwinowski, Schevitz *et al.* 1988). Thus the contribution of indirect readout to the ability of a protein to bind DNA was early suspected to be important; however, the chemical component of binding also remained to be understood.

Between 1980 and 1999, more than 200 protein/DNA structures were documented as individual structures and as binding motifs, highlighting the important role of hydrogen bonds and non-polar interactions. (Jones, van Heyningen et al. 1999) Local structure and deformability of DNA were known to be a function of base sequence due to structural work exemplifying two major modes of induced DNA bending by a protein.

One mode was a local, severe bend such as that seen in the catabolite activator protein (CAP) /DNA complex, where major groove contacts result in moderately high roll angles in three consecutive basepairs. (El Hassan and Calladine 1998) The CAP protein is an  $\alpha/\beta$  structure and activates transcription by binding to a DNA site located in or upstream of the core promoter and interacting with the RNA polymerase  $\alpha$  subunit (figure 1.03). When CAP binds DNA, it introduces a sharp kink, characterized by a roll angle of  $\sim 40^\circ$  and a twist angle of  $\sim 20^\circ$  between positions Thy6 and Gua7 in the DNA site. (Parkinson, Wilson et al. 1996) Substitution of the pyrimidine Thy6 with another pyrimidine cytosine had little effect on the global DNA geometry of the complex, but substitution with adenine or guanine purines decreased roll angles to  $\sim 20^\circ$ , and twist angles to  $\sim 17^\circ$  in the crystal structures, indicating that the flexible pyrimidine-purine central step was critical energetically for proper complex formation. (Chen, Vojtechovsky et al. 2001; Napoli, Lawson et al. 2006)

Similarly, the TATA-binding protein (TBP) induces severe bending in its cognate DNA. TBP is an  $\alpha/\beta$  structure that interacts with the minor groove of the DNA via a long  $\beta$  sheet lying in the minor groove of the untwisted DNA (figure 1.04). It is only as a result

of minor groove interactions untwisting the DNA that there is distortion of seven DNA basepairs in the cognate sequence. The untwisting results in particularly high roll angles and low twist, effecting a 90 degree bend ultimately. (Juo, Chiu et al. 1996)

These proteins create unnaturally high roll angles, or otherwise distort the DNA in such a way that the energy needed is substantial, but it is also possible for a protein to manipulate the natural flexibility of the DNA in particular ways to effect specificity. In naked DNA, natural variations in roll angles would normally cancel out, but proteins can create in phase manipulation of roll angles within the periodic repeat of the double helix, and thereby create a gentle bend. (El Hassan and Calladine 1998)

The 434 repressor creates a gentle bend in the DNA within the range of nucleotide step distortion seen in naked DNA structures (Lilley 1986; Hagerman 1990; Dickerson, Goodsell et al. 1996), and facilitating bending around the histone octamer core. In complexes with bacteriophage 434 binding sites, the 434 repressor does not contact the central 4 basepairs of the 14 basepair site. Operators with AT or TA basepairs at these positions bind repressor more strongly than those bearing CG or GC, suggesting that these bases are important for the repressor's ability to discriminate between operators. (Koudelka and Carlson 1992)

Experiments showed that there was a relationship between the intrinsic twist of an operator, as determined by sequence, and its affinity for repressor; an operator with a lower affinity is undertwisted relative to an operator with higher affinity. (Koudelka and

Carlson 1992) Further changes in these central four basepairs altered the binding site affinity for the repressor. Specifically, a single base insertion mimicking the natural binding site being underwound and allowing for increased twist in complex, is still able to be bound by the repressor, but a central base deletion representing overwinding of the DNA is not able to be bound. (Koudelka 1998) These results are consistent with the hypothesis that the sequence of the central basepairs allows for overtwisting of the DNA double helix when in complex with the repressor.

Surprisingly, however, the number of hydrogen bonds in GC basepairs vs. AT basepairs was revealed to have no role in determining the relative affinity of a DNA site for repressor in the central four basepairs. Rather, the defining characteristic was the presence or absence of the N2-NH2 group on the purine bases at the binding site center. The N2-NH2 group on bases at the center of the 434 binding site appears to destabilize the repressor/DNA complexes by decreasing the intimacy of the specific repressor/DNA contacts and increasing the reliance on protein contacts to the backbone to compensate. (Mauro, Pawlowski et al. 2003)

Another protein/DNA complex which has prompted substantial analysis of the role of indirect readout is the methionine repressor (MetJ), a homodimer of 104 amino acid subunits, each with three alpha helices and one beta strand which form a short sheet in the dimerization interface (figure 1.05). Corepressor (s-adenosyl methionine, or SAM) binding pockets are symmetric and the dimerization anti-parallel beta sheet also functions as the DNA binding domain, with the sheet making contacts in the major groove of the

metbox site.(Somers and Phillips 1992) MetJ represses the transcription of genes involved in methionine biosynthesis by binding to 2-5 copies of the metbox/gene, with affinity for the metbox varying with deviations from consensus. The affinity variations are interesting because MetJ seems very sensitive to particular base changes, including positions not directly contacted by the protein, but some loss of affinity can be overcome with higher copy number of the metbox in the promotor region. (Phillips, Manfield et al. 1989)

Crystallography experiments testing the structure of two MetJ repressors bound to a double metbox site tested variations in bases thought to be read indirectly between the metboxes themselves (figure 4). Results revealed small compensatory variations in the sugar-phosphate backbone conformation and some direct contacts. The basestep at the center of the space between metboxes displays a bend towards the major groove, with flanking three basesteps showing concurrent helical twist and narrowing of the minor groove. This structure was somewhat disrupted in the reversal of the TA step for AT at the central bases, suggesting that the functional decrease in affinity (75 fold) might be due to decreased flexibility at that step influencing cooperativity at the multiple metboxes.(Garvie and Phillips 2000)

### **The Purine Repressor (PurR) : *E. coli* master regulator of purine metabolism.**

In *E. coli*, purine nucleotides can be derived by salvage pathways from exogenous purine pools, or they can be synthesized *de novo*, as with the eleven enzyme mediated steps

needed to produce inosine monophosphate, the precursor to adenosine or guanosine monophosphate; the transcription of these eleven enzymes in nine separate loci is regulated by the Purine Repressor (PurR, figure 1.06a); the first step in this cascade is accomplished by an enzyme called the glutamine PRPP (5-phospho—D-ribosyl-1-pyrophosphate) amidotransferase, encoded by the *purF* gene. The *purF* gene, possibly due to its status as the first enzyme gene in a highly regulated pathway, is the most stringently repressed (~17 fold) by PurR in the presence of a corepressor purine (hypoxanthine or xanthine) binding a 16 base pair operator situated over the -35 element (figure 1.06c); *purF* is also the best characterized operator. The rest of the genes in the pathway are downregulated by at least 10 fold; they are joined in the operon by genes encoding enzymes to convert IMP to GMP and AMP, although they are repressed only 2-5 fold. The differences in repression reflect an allowance for these genes to be transcribed as part of the salvage pathways even when the *de novo* IMP synthesis is stringently repressed; this is likely structurally accomplished by variations in the sequence of the operator as well as location in the gene landscape. (Figure 1.06c) PurR additionally autoregulates itself (2-3 fold repression), and purine regulated genes involved in nucleotide metabolism, making up an operon of 21 known genes. (Figure 1.06b)

The structure of the Purine Repressor (PurR) was first published in 1994 (figure 1.01) and represented the first high-resolution full-length structure of a LacI family member (Schumacher, Choi et al. 1994). Previous structural work within the family being limited to NMR structures of the LacI DNA binding domain only, and with limited information:

the binding motif was found to be a helix turn helix (figure 1.07), but the full source of DNA sequence discrimination was incomplete, and the understanding of dimerization, signal transduction or effector domain structure could not be addressed by these structures.

The LacI family is a family of transcriptional repressors with highly homologous primary structures. Members of this family function by binding similar pseudopalendromic operator sites generally 16-18 base pairs long. Structurally, LacI family members have two domains; the first is a large C-terminal effector domain, generally 250 residues or more, and the second is the N-terminal DNA binding domain, about 60 residues. Most LacI family members bind DNA with high affinity in the absence of their effector molecules, but PurR does not; PurR is also unusual in having two corepressors, hypoxanthine or guanine. (Figure 1.01, (Schumacher, Glasfeld et al. 1997))

PurR is a dimeric protein of 341 residues/subunit that binds any of 21 known genes. In order to act as a transcriptional repressor, PurR requires one of two purine corepressors, hypoxanthine or guanine, to bind specifically to a 16 base pair pseudo-palendromic operator site, which allows the 21 operators to be described as 42 half-sites. In the crystal structure, the operator *purF* was used, having this sequence and numbering:

```
1 2 3 4 5 6 7 8 9 9'8'7'6'5'4'3'2'  
A A A G A A A A C G T T T G C G T  
T T C T T T T G C A A A C G C A T  
2'3'4'5'6'7'8'9'9 8 7 6 5 4 3 2 1
```

The structure of PurR is bipartite: The N-terminal DNA Binding Domain (DBD, residues 1-60) contains a classic helix-turn-helix (HTH) binding motif followed by a loop and additional helix. Both direct and water mediated electrostatic contacts are made, and the three helix bundle of the DNA binding motif is responsible for major groove specificity. The trailing loop and 4<sup>th</sup> helix connecting the domain to the corepressor binding domain (the hinge helix), along with the dimerization mates, make additional specific contacts by inserting into the minor groove and kinking the purF operator by nearly 45 degrees via the interdigitation of residues Leu54 and its symmetry mate within the central CpG step. There are also flanking contacts within the minor groove, notably by Lys55. (Schumacher, Choi et al. 1994)

The C-terminal corepressor binding domain comprises residues 61-341 and is responsible for dimerization as well as corepressor specificity and binding. Within the CBD are two topologically similar  $\alpha/\beta$  subdomains (N- and C-terminal subdomains) with three crossover connections. The corepressor molecule is bound in the cleft between these subdomains using polar, non-polar and aromatic interactions. Specifically, there are direct and water mediated contacts to Tyr73, Phe74, Arg190, Thr192, Phe221, Asp275 and Arg196. (Schumacher, Glasfeld et al. 1997)

Corepressor binding, however, while required for high affinity binding, occurs more than 40 Å away from the DNA. The initial and later structures of PurR suggested that when hypoxanthine or guanine bound to PurR (signaling an excess of purines and acting as an

environmental switch to stop *de novo* synthesis) PurR would be activated to bind DNA by repositioning the hinge region such that they interact, and undergo a coil-to-helix transition, allowing those residues to bind in the minor groove. The conformational change required to go from the unbound, open form (crystallized as the CBD alone) to the corepressor bound closed form involves a 17-23° hinge bending rotation between the CBD subdomains (notable Tyr73, Trp147, Asp160) in order to allow for the correct positioning for hinge helix formation. (Schumacher, Choi et al. 1994; Schumacher, Choi et al. 1995)

DNA binding specificity in PurR is dependent upon base specific contacts made by the HTH motif in the major groove and DNA deformability and contacts made to the minor groove by the hinge helix (figure 1.09). The interdigitation in the minor groove by the dyad related Leu54 leads to a 49° kink at the central CpG step, broadening the minor groove and locally unwinding the DNA. (Schumacher, Choi et al. 1994) This central kink is supported by contacts between Lys55 C $\epsilon$  and the C2 of Ade8, and between the N $\zeta$  and the N3 of Ade8 or the O2 of Thy7'. Because the B-factors for C $\delta$ , C $\epsilon$ , and N $\zeta$  are all high (at least 80, please see the methods chapter for a discussion of B-factors), this indicates sufficient flexibility that any of these contacts might predominate and exchange and combine to give an averaged contribution to binding. To further examine the role of the specific base contacts made in the minor groove, the mutation of Lys55 to Ala was created and found to have no impact on the globular structure of the protein or DNA, while having a substantial impact on binding affinity (decreasing affinity by 320 fold); it was theorized that the loss of Lys55 contacts accounted for the change in affinity.

Although the possibility existed that the mutation would expose the area to solvent, no solvent molecules were seen nearby in the 2.7Å structure, and the local parameters for the Ade8:Thy8' base pair did not show substantial distortion, suggesting that the major role of Lys55 is to enhance the affinity of the repressor for the operator. (Schumacher, Choi et al. 1994; Schumacher, Choi et al. 1995; Glasfeld, Koehler et al. 1999)

To further elucidate the role of the deformability of the DNA and the role of Lys55 near the structure of the central kink of the DNA, Ade8 and it's base pair (the primary base pair contacted by Lys55) were substituted for cytosine, thymine or guanine (Cyt8, Thy8, Gua8, respectively) and the structures and function (binding affinity) were examined. Despite escalating dissociation constants (PurR had a 4 fold decrease in affinity for Thy8 containing operator, and 14 fold decrease for Cyt8 DNA, and regardless of whether the wild type or Lys55Ala protein was used,) the global structure of the DNA was unperturbed in the crystal structures (Glasfeld, Koehler et al. 1999). The role of the deformability of the DNA remained unclear.

A number of studies have been undertaken that explore the deformability of DNA out of the context of a bound protein complex, however; the Dickerson laboratory has demonstrated that purine-pyrimidine steps, and most notably CG steps, are the most easily bent, probably due to electrostatic repulsion in the functional groups as well as conflicting propeller twist in the base pairs combining to effect a very poor ability to stack tightly as neighboring base pairs. Conversely, purine tracts, and specifically adenine tracts (Nelson, Finch et al. 1987), stack very well together as their high degree of

identical propellor twisting allows them to stack much like puzzle pieces; so much so that they have been termed 'rigid rods' if two or more adenines neighbor each other. (Figure 1.09, (Sauer 1995; el Hassan and Calladine 1996; el Hassan and Calladine 1996; Dickerson and Chiu 1997; Dickerson 1998; Lavery and Lebrun 1999; Lebrun and Lavery 1999; Garvie and Wolberger 2001; Bosch, Campillo et al. 2003; Kalodimos, Biris et al. 2004))

PurR's manipulation of DNA on binding has characteristics notable in light of these studies, including a central CG step where intercalation of Lys54 creates a dramatic 55° kink in the DNA, and flanking A-tracts to either side, which are highly conserved within the operon, but undercontacted from a purely direct readout standpoint (figure 1.10)

### **The TetR Family of Transcriptional Regulators**

Prokaryotic transcriptional regulators are classified in families (such as the LacI family or the TetR family) on the basis of sequence similarity and structural and functional criteria. TetR family members are identified through sequence similarity in the HTH DNA-binding domain control genes involved in multidrug resistance, catabolic pathways, antibiotic biosynthesis, osmotic stress and pathogenicity. The TetR family currently comprises some 73 members according to the profile established by Ramos et al. based on TetR and QacR (Aramaki, Yagi et al. 1995; Ramos, Martinez-Bueno et al. 2005).

The conserved DNA binding motif that is the identifier for the family comprises not only the HTH motif, which is a common DNA binding motif for other prokaryotic transcriptional regulators as well (including PurR), but also a stretch of conserved

residues which in the QacR and TetR structures corresponds to the majority of  $\alpha$ -helix 1, the HTH motif comprised of  $\alpha$ 2 and  $\alpha$ 3, and 5 residues of  $\alpha$ -helix 4 that connect the DNA-binding region to the core of the protein (figure 1.12). Such a high degree of homology (averaging approximately 60%) and identity (~33%) in the DNA binding domain suggests a conserved structure; this theory is supported by the three-dimensional structures of TetR (figure 1.13), QacR (figure 1.14), CprB and EthR (figure 1.15) in this region. (Orth, Schnappinger et al. 2000; Schumacher, Miller et al. 2001; Schumacher, Miller et al. 2002; Schumacher and Brennan 2003; Dover, Corsino et al. 2004; Engohang-Ndong, Baillat et al. 2004; Frenois, Engohang-Ndong et al. 2004; Natsume, Ohnishi et al. 2004) As expected, there is little sequence conservation outside the DNA binding domain, reflecting the differences in signal sensed by the various regulators in the family, although there is homology in the secondary structural elements and their placement. (Schumacher, Miller et al. 2001; Ramos, Martinez-Bueno et al. 2005)

TetR, a transcriptional repressor and the protein for which the family was named, in the absence of signal (tetracycline, one of the most commonly used antibiotics) binds to DNA and prevents the transcription of TetA, a membrane bound pump, and of itself; although the gene for TetR is differently oriented from TetA, they have identical 15 base pair pseudopalendromic operators, spaced 11 base pairs apart and overlapping the promoters for the genes under regulation. (Unger, Klock et al. 1984; Orth, Schnappinger et al. 2000; Ramos, Martinez-Bueno et al. 2005)

TetR is a homodimer in both the DNA and drug bound forms. (Unger, Klock et al. 1984; Orth, Schnappinger et al. 1999; Orth, Schnappinger et al. 2000) The global structure for each monomer includes 10  $\alpha$  helices along with assorted loops and turns. (figure 1.13) Helices 1, 2 and 3 are involved in DNA binding, where  $\alpha$ 4 is a connector helix and  $\alpha$ 5,  $\alpha$ 6,  $\alpha$ 7,  $\alpha$ 8,  $\alpha$ 9 and  $\alpha$ 10 form the regulatory and dimerization domain. Tetracycline enters the binding pocket through an entrance to the cavity formed by the dimer's  $\alpha$ 8' and  $\alpha$ 9' helices. Once there, the first ring, A, of the drug contacts loop 4-5 at the back of the pocket while the complexed magnesium ion mediates the contact between center rings B and C of the drug and TetR's His 100 and Thr103 of  $\alpha$ 6, displacing the helix and instigating a conformational change of Arg104 and Pro105 to  $\beta$  turn; this in turn displaces  $\alpha$ 4 and  $\alpha$ 3 sequentially into the DNA binding domain. The shift in position of  $\alpha$ 3 allows the dissociation of TetR from the DNA. (Orth, Schnappinger et al. 2000)

TetR in the absence of drug binds its cognate DNA as a homodimer in two successive major grooves, each DNA binding domain contacting 6 base pairs; monomer A contacting bases -4 through -7 of the main strand and +4 to +2 of the complimentary strand, and monomer A' the reverse on the main and complimentary strands. There are no water mediated contacts in the interface, as all the crucial interactions are hydrophobic, and the stability of the DNA binding domains themselves are due to a hydrophobic core interaction in the three helix bundles (comprising  $\alpha$ 1, 2 and 3).

Helix  $\alpha$ 3 (Gln38 – His44) is the main recognition element for sequence specificity and all the residues in the helix contribute to DNA sequence recognition with the exception of

Leu41, which interacts with the hydrophobic core of the three-helix bundle. Specifically, Thr40 interacts with Thy-7 and Cyt-6 of the main strand, and Trp43 also contacts Thy-7. Pro 39 makes contact with Thy-5, Ade-4 (main strand) and Thy+4 of the complementary strand; Thy+4 is additionally contacted by Tyr42 and its neighbor, Thy+3 interacts with Gln38. An additional specific contact comes from Arg28 in  $\alpha 2$ , which contacts Gua+2 of the complementary strand, and there is also a contact to the DNA region from outside the HTH motif in  $\alpha 4$ , as Lys48 makes a non-specific contact.

Upon binding DNA, the recognition helix,  $\alpha 3$ , undergoes conformational change in the N-terminal region of the helix to form a  $3_{10}$  helical turn. The DNA-protein hydrogen bonding at Arg28-Gua+2 and Gln38-Ade+3 increase the separation between the first and second base pairs of the DNA from 3.4 to 3.8 Å. The flanking phosphate groups around Gua+2 make contacts to Thr26, Thr27, Tyr42 and Lys48 which effects a kinking at Gua+2 away from TetR; base pairs +3 to +6 compensate for the kink by bending toward the DNA. (figure 1.13b)

QacR is the only other member of the TetR family for which we have crystal structures for both drug bound (for multiple drugs) and DNA bound forms (figure 1.14). QacR is the transcriptional regulator for the multidrug transporter gene *qacA*, a pump that confers resistance to mono and bivalent cationic lipophilic antiseptics and disinfectants such as the quaternary ammonium compounds for which it was named. Both *qacA* and *qacR* are part of the *qac* locus and are plasmid encoded, although divergently transcribed. When not bound to drug, the 188 residue QacR protein binds two nested palindromes

downstream from the *qacA* promoter and overlapping its transcription start site to repress the transcription (apparently by hindering the transition of RNA polymerase to a productively transcribing state rather than by blocking binding outright.)

Like the QacA pump, QacR binds cationic lipophilic drugs such as rhodamine 6G, crystal violet and ethidium as well as some bivalent cationic dyes and plant alkaloids, and dissociates from the DNA to allow transcription of the efflux pump. Equilibrium dialysis and isothermal titration calorimetry studies of the drug bound form indicate that only one monomer of the QacR dimer binds the drug, in a 2:1 stoichiometry, so that only one binding pocket appears to be available to bind drug.

Similar to TetR, crystal structures of QacR show that it is an all-helical protein and a functional dimer containing a DNA binding HTH motif within an N-terminal three helix bundle, establishing its homology with the TetR family. The stoichiometry of DNA binding differs notably from TetR, however, in that QacR binds its operator with two dimers rather than just one (figure 1.14a). The crystal structure additionally shows an expansive and multifaceted drug pocket comprising nearly  $1,100 \text{ \AA}^2$  and multiple overlapping subpockets (figure 1.14b); reflecting the difference between the multidrug binding properties of QacR as opposed to the single target, tetracycline, of TetR and similar to that reported for AcrB, another multidrug binding protein.

The drug binding domain of QacR consists of 6  $\alpha$ -helices. The mostly buried pocket has its entrance formed by the opening between  $\alpha 6$ ,  $\alpha 7$ ,  $\alpha 8$  and  $\alpha 8'$ , asymmetrically and

related to the single drug/dimer stoichiometry. During the drug-binding, DNA dissociation process, the drug bound monomer undergoes major structural changes: a coil to helix transition in residues 89-93 extends  $\alpha 5$  by a turn and causes the expulsion of Tyr92 and 93 from the hydrophobic core and the binding pocket where they had been acting as drug surrogates. This also relocates  $\alpha 6$ , which is connected to the DNA binding domain; the DBD is shifted 9Å, rotated 37°, and is no longer optimally spaced from the other DBD, leading ultimately to dissociation from the operator.

Overall, the general topology of QacR is similar to TetR. The proteins' secondary structural elements overlay nicely even in the less homologous C-terminal domain; the only difference is that QacR contributes helices  $\alpha 8$  and  $\alpha 9$ , and TetR contributes  $\alpha 8$  and  $\alpha 10$  to the 4 helix bundle dimerization interface (the area between  $\alpha 8$  and  $\alpha 9$  in the crystal structure is disordered for TetR, interestingly). The DNA binding domains, with their HTH motifs embedded in three helix bundles and high homology between family members (figure 1.12), have expectedly similar secondary structure homology; both QacR and TetR also bind partial palindromes, and in this sense, the interactions of the protein with the target sequences are equivalent, despite the stoichiometric binding differences, and may be representative of a trend in the family. QacR, though, binds two overlapping partial palindromes within the same fragment and, despite the identical symmetric bases in the sequence, only one dimer binds symmetrically in a palindrome, the other partially overlaps the sequence bound by the first, likely due to the cooperative nature of the two dimers binding mechanisms.

The QacR dimers are perhaps best designated the A and B monomers (see figure 1.14a), with the sublabeling of those monomers proximal and distal to the center of the operator. In both the A and B distal monomers,  $\alpha 3$  makes the most extensive specific interactions; In A distal, Tyr41 makes a hydrophobic contact to the DNA mainstrand at Thy-10 and the phosphate of position -11, while Tyr40 contacts Thy+7 of the complimentary strand. Specific hydrogen bonds are made between Lys36 and Gua+6 of the complimentary strand and Gly37 and Gua-8 of the main strand, a key interaction as Gua-8 is also *qacA*'s transcriptional start site. In the proximal monomers, there are extensive contacts as well, some critical interactions being between Tyr44 and Cyt-6 (main strand) and Tyr40 and Thy+3 (complimentary strand) in the B proximal monomer; Gly37 and Gua-4 (complimentary strand) and Lys36 and Gua+1 (main strand). Additionally, the proximal monomers make phosphate backbone contacts with helix  $\alpha 2$ , loop 2-3, helix  $\alpha 3$  and the N-terminal dipole moment from helix  $\alpha 1$ .

Despite extensive contacts to the DNA at overlapping sites, however, the A and B dimers do not get within a 5Å distance of each other and instead bind major grooves on nearly opposite faces of the DNA; the cooperative nature of the two dimers' binding appears to come not from protein-protein interactions, but from the favorable underwinding of B-DNA at the overlapping site. This transition from B-DNA to the high-affinity undertwisted conformation allows the distance between dimeric HTH motifs to be optimally spaced at 37Å rather than what would be 34Å in the more compact B-DNA.

QacR's fairly even widening (the maximum bend is 3Å at most, although the variation is reflected in different center to center distance measurements for the dimers) of the major grooves along the entire binding site is in contrast to TetR, which kinks its binding site in a single 17° bend toward the protein to achieve optimal HTH separation. Additionally, TetR uses Arg26, a residue outside the  $\alpha 3$  recognition helix, to make a base pair specific contact where QacR's base specific contacts are restricted to  $\alpha 3$ . These different mechanisms of binding still achieve a similar degree of specificity, however. It is also worth noting that they represent two different groups within the subfamily of members involved in resistance to toxic substances: those that bind a broad range of structurally diverse ligands (QacR) and those that bind with high specificity to very few ligands (TetR).

QacR may be more representative of the family in general, however, when it comes to changes in stoichiometry with binding; studies of EthR, a transcriptional regulator of the monooxygenase EthA which influences the activation of the cancer therapeutic ethionomide and is directly related to resistance, indicate that EthR also binds operator DNA with a different stoichiometry than it binds drug. Specifically, while crystal structures indicate that EthR is the expected all-helical homodimer when bound to ethionomide, surface plasmon resonance studies suggest it is an octamer when bound to its 55 base pair operator (part of the 75 base pair intergenic region between *ethR* and *ethA* (Baulard, Betts et al. 2000; Dover, Corsino et al. 2004; Engohang-Ndong, Baillat et al. 2004; Frenois, Engohang-Ndong et al. 2004)); TetR, by comparison, binds a 15 base pair operator and QacR contacts 22 base pairs. Additionally, the  $\gamma$ -butyrolactone

autoregulatory factor receptor (CprB) from *Streptomyces*, related to the A-factor receptor protein identified as an essential component of streptomycin resistance, has been crystallized in apo form (the ligand has yet to be identified) and the structure solved. The structure of homodimeric apo CprB so closely resembles QacR bound to DNA (except for the lack of a 10<sup>th</sup>  $\alpha$ -helix) that the authors were able to superimpose the DNA-binding domains with an RMSD of 1.48Å over 71 backbone carbon atoms, and predict the core residues of the domain's three helix bundle (Ile14, Ile15, Ala18, Phe22, Leu32, Ile35, Leu46 and Phe50 (Schumacher, Miller et al. 2002; Natsume, Ohnishi et al. 2004)).

### **The Multiple Transferable Resistance Regulator (MtrR) from *Neisseria gonorrhoeae*.**

MtrR also belongs to the TetR family functional subgroup involved in regulation of efflux pumps and transporters involved in antibiotic resistance and tolerance to toxic chemicals (Martinez-Bueno, Molina-Henares et al. 2004). MtrR is a transcriptional repressor that regulates transcription of the *mtrCDE* tandem gene encoding a multidrug efflux pump, and as part of a more complex circuit either directly or indirectly regulates FarAB expression, another efflux pump that utilizes the MtrE outer membrane channel protein as part of its system. (Hagman, Pan et al. 1995; Hagman and Shafer 1995; Lucas 1997; Lee, Rouquette-Loughlin et al. 2003; Hoffmann, Williams et al. 2005)

In the MtrCDE efflux pump MtrE is linked to the MtrD multidrug efflux transporter protein that belongs to the resistance/nodulation/division transporter family by MtrC, which is a membrane fusion protein; all of which together recognize and efflux diverse antibacterial hydrophobic

agents (HAs) and peptides to the extracellular environment (Hagman, Pan et al. 1995; Delahay, Robertson et al. 1997; Hagman, Lucas et al. 1997; Veal and Shafer 2003). Repressed expression of MtrR, a divergently transcribed gene (although the -35 box overlaps with that of *mtrC*, figure 1.16) allows high expression of the MtrCDE efflux pump and concomitant resistance to HAs (Hagman and Shafer 1995; Lucas 1997). Similarly, hypersensitivity in strains of *N. gonorrhoeae* can be traced back to mutations in *mtrCDE* (Guymon, Walstad et al. 1978; Hagman, Pan et al. 1995; Veal, Yellen et al. 1998).

The *mtrR* gene encodes a 210 amino acid residue, ~23 kDa protein (MtrR). MtrR contains a putative N-terminal helix-turn-helix (HTH) motif and amino acid sequence similarity to several members of the TetR family, but it bears the strongest resemblance to AcrR: 53% identity, 78% homology (figure 1.12.) Footprinting experiments have shown MtrR binds to a 40 base pair region between *mtrR*'s -10 and -35 boxes containing an inverted repeat, though further pinpointing the binding site indicates a 31 base pair imperfect direct repeat (Hagman and Shafer 1995; Lucas 1997). Alignments have indicated many critical residues in the recognition helix identified in TetR and QacR structures are conserved in MtrR, notably the Tyr-Trp-His motif (Tyr-Tyr-His in QacR) at TetR positions 37-39 which make contacts in to Thymine bases and the phosphate backbone in the TetR structure and Lys43, a critical residue thought to adjust the HTH motif into position for binding, conserved in over 77% of the members of the TetR family. Other residues at positions critical for recognition but less stringently conserved (or perhaps related to specificity for a given promotor) are well conserved in over 20% of the family and consistent with hydrophobic or polar requirements for the position (Aramaki, Yagi et al. 1995; Orth, Schnappinger et al. 2000; Schumacher, Miller et al. 2002; Ramos, Martinez-Bueno et al.

2005). Additionally, a number of basic residues not present in QacR or TetR (or visible in the structures as a result) are present at the N-terminus of MtrR which would, if modeled on the QacR HTH domain, be in an ideal position to contact the acidic DNA (figure 1.17); a hypothesis which becomes even more attractive when one notes that the proposed binding site is potentially large enough (up to 31 base pairs according to the footprinting assay (Lucas 1997)) to accommodate extra contacts.

The circuit of regulation in the *mtr* system is becoming more complex, however, as further investigation in the field reveals the AraC-like MtrA, which may be a second component activator of the *mtr* system (Rouquette, Harmon et al. 1999). Also recently identified is MtrF, encoded by a gene downstream of *mtrR* and a putative cytoplasmic membrane protein under MtrR regulation and possibly involved in high level detergent resistance in conjunction with MtrCDE (Rouquette, Harmon et al. 1999; Folster and Shafer 2005). The precise role of MtrR in the regulation of the *farAB* genes, the fatty acid resistance pump proteins, is also unclear. Although some degree of control is evident from knock-out studies and similar inverted repeats (figure 1.16) can be found in the promoter region of both *farAB* and *MtrCDE* genes, direct regulation has yet to be shown, and the role of the transcriptional regulator FarR is also unclear (Lee and Shafer 1999; Lee, Rouquette-Loughlin et al. 2003). Proof of ligand binding to MtrR also remains elusive, even though the efflux pump binds HAs.

The first MtrR paper (chapter 4), which has already been published in the July issue of the *Journal of Bacteriology*, was a primary characterization of MtrR, including it's secondary structural characteristics compared to QacR, an exploration of the minimum

binding site as well as the optimum site requirements, binding constants and stoichiometry experiments.

The second MtrR-related paper (chapter 5) is a more detailed exploration of its DNA binding properties. The experiments for this paper include multiple permutations of the DNA binding site, exploration of the binding affinity for the IR sites vs. the DR site, G tract substitution for the flanking A tracts, and half site vs. full site experiments.

Additionally, we have created deletion and site mutants (in an alanine scan of basic residues) within the basic 8 amino-acid N-terminal region of the DNA binding domain in order to describe thoroughly the influence of this extension of the DNA binding domain on DNA binding.

The remaining paper in this thesis (chapter 3) explores the structural and functional effect of indirect readout on the binding of PurR to its binding site using binding studies in tandem with x-ray crystallography to document the effect of a mutation within the binding site on the structure of the protein.

The statement of purpose for the thesis: Explore the nature of protein-DNA interactions by studying two bacterial model systems: the Purine Repressor (PurR) from *Escherichia coli* and the Multiple Transferable Resistance Regulator (MtrR) from *Neisseria gonorrhoeae*. In the well characterized PurR system, we use x-ray crystallography to explore the structural changes inherent in indirect readout in connection with the functional implications of altering a base not directly contacted by the protein. With MtrR

we shall see the initial biochemical and biophysical characterization of a protein, followed by a more detailed exploration of the affinity of the protein for DNA and the contribution to the electrostatic contacts by the N-terminal basic region.

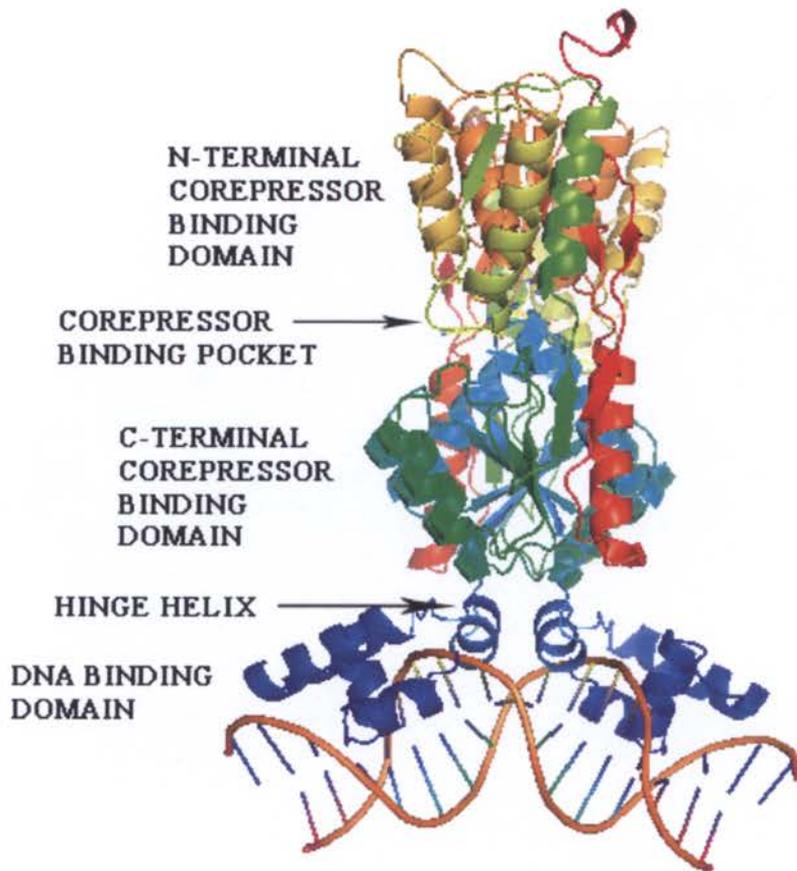


Figure 1.01: Cartoon model of the dimeric PurR/hypoxanthine/*purF* complex (PDB code 1QPZ) showing the corepressor and DNA binding domains, as well as the corepressor binding domains and hinge helix.

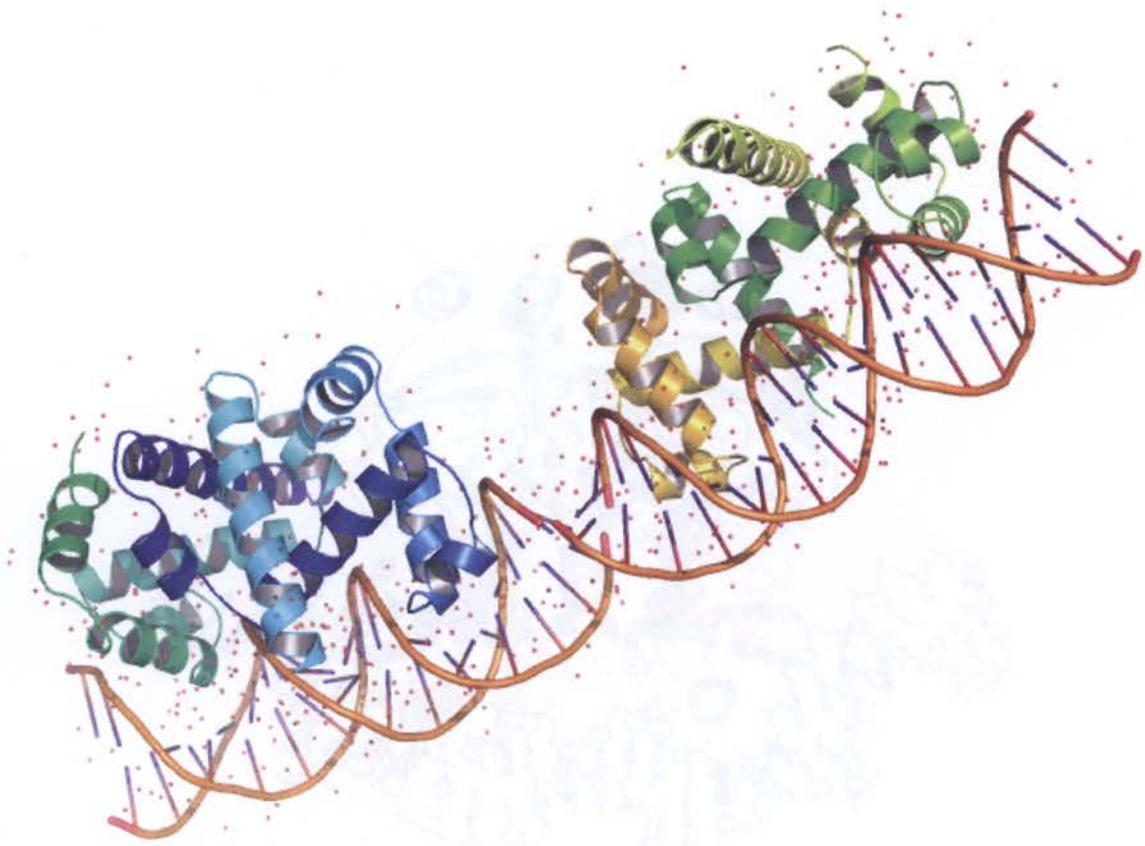


Figure 1.02: The Tryptophan Repressor (TrpR)-operator complex revealed an extensive contact surface including 24 direct van der Waals contacts between the  $\alpha$ -helical protein and DNA, and 6 solvent mediated hydrogen bonds to the phosphates of the DNA backbone. The first transcriptional regulator to be crystallized and solved in complex with its cognate DNA to high resolution, there were, surprisingly, no direct hydrogen bonds or non-polar contacts to the bases that explained the specificity of the protein for the operator. Structure is 1TRO in the Protein Data Bank.

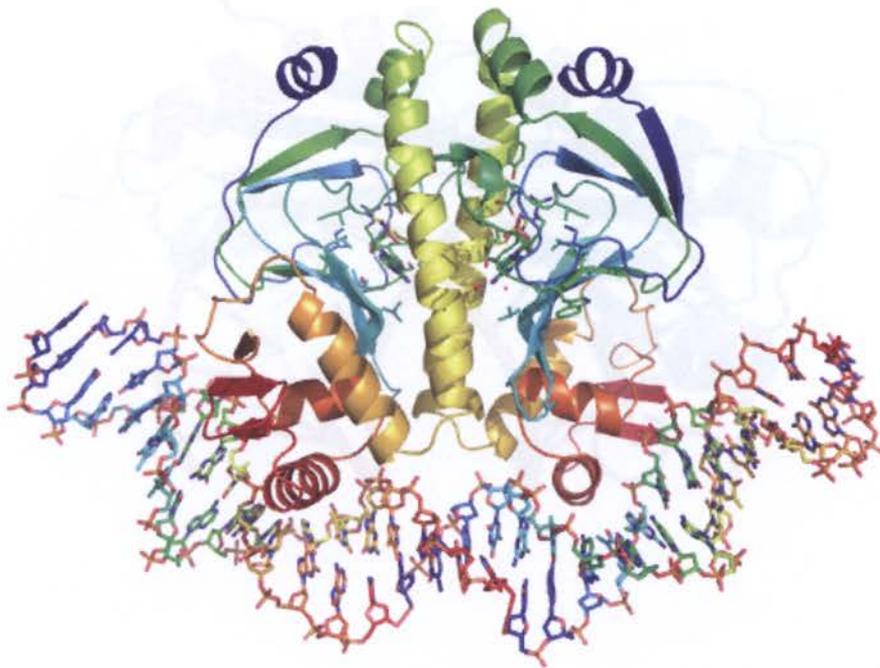


Figure 1.03: The CAP-DNA complex. CAP binds to the DNA through a sharp kink in the DNA, which is characterized by a roll angle of  $\sim 40^\circ$  and a twist angle of  $\sim 20^\circ$  between the central Thymine and Guanine in the DNA site, a basepair step that has been shown energetically to require the flexibility of a purine-pyrimidine step for proper complex formation. Structure is 1J59 in the Protein Data Bank.

Figure 1.03: The Catabolite Activator Protein (CAP)-DNA complex. When CAP binds DNA, it introduces a sharp kink, characterized by a roll angle of  $\sim 40^\circ$  and a twist angle of  $\sim 20^\circ$  between the central Thymine and Guanine in the DNA site, a basepair step that has been shown energetically to require the flexibility of a purine-pyrimidine step for proper complex formation. Structure is 1J59 in the Protein Data Bank.

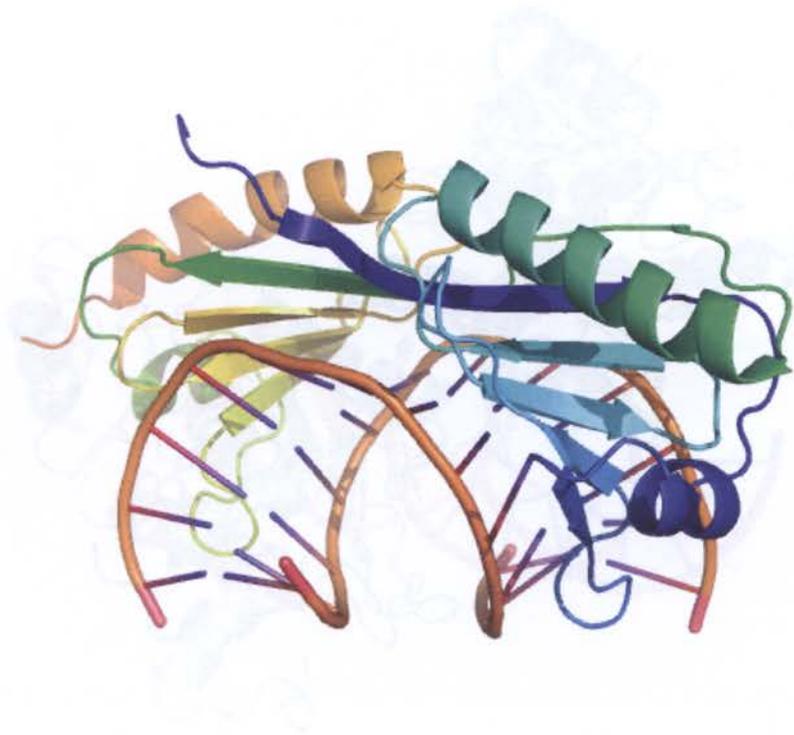


Figure 1.04: The TATA-binding protein (TBP) -operator complex. TBP introduces a 90° bend in the DNA through minor groove interactions, distortion of seven basepair steps and untwisting of the DNA resulting in high roll angles for the basepairs. Structure is 1TGH in the Protein Data Bank.

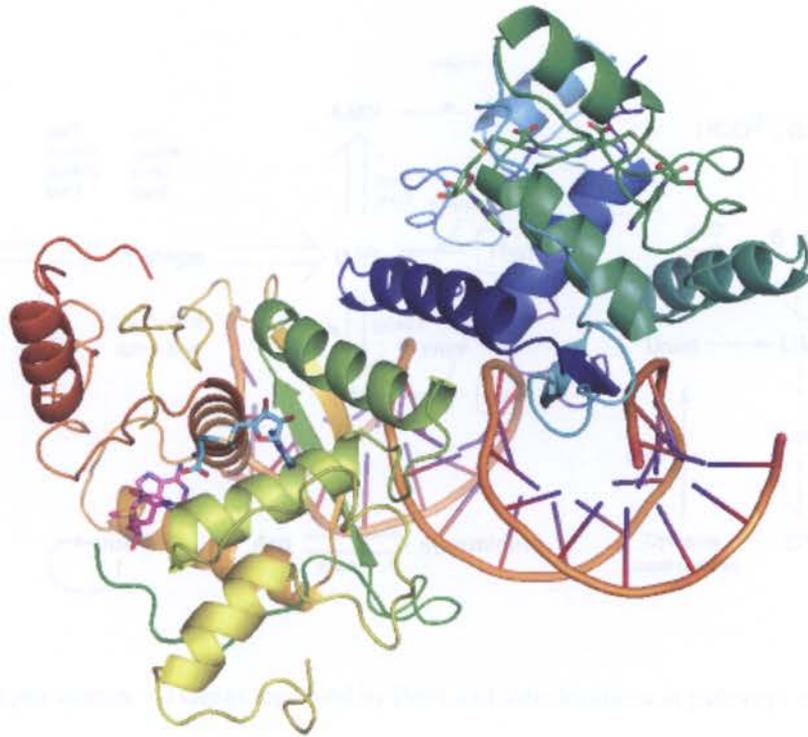


Figure 1.05: Two oligomers of the methionine repressor (MetJ) bound to two operator metboxes. The number consecutive metboxes modulate transcriptional control along with fidelity to the consensus metbox sequence. The area between metboxes is not directly contacted by the proteins, but is indirectly read as flexibility is thought to influence the ability of multiple MetJ dimers to bind cooperatively. PDB accession number: 1MJ2

A

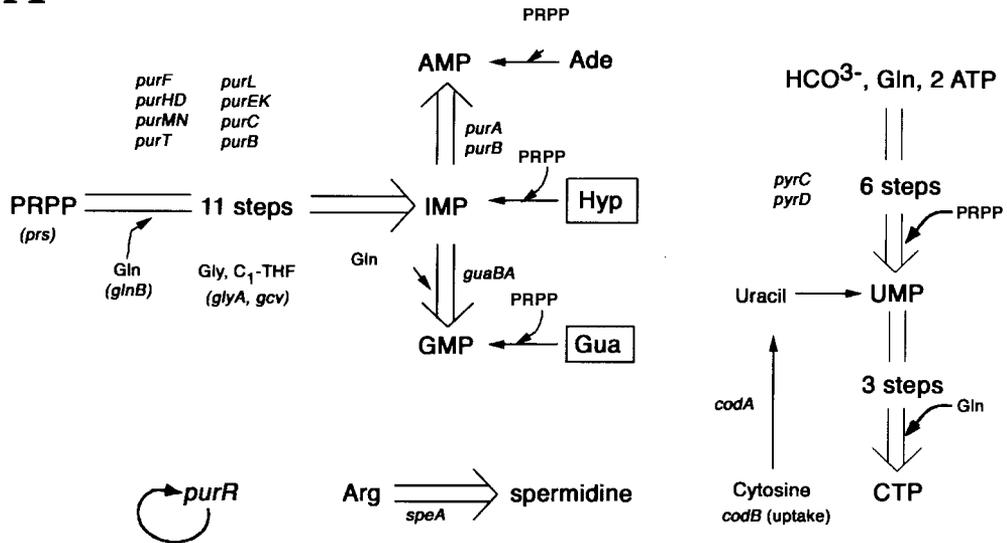
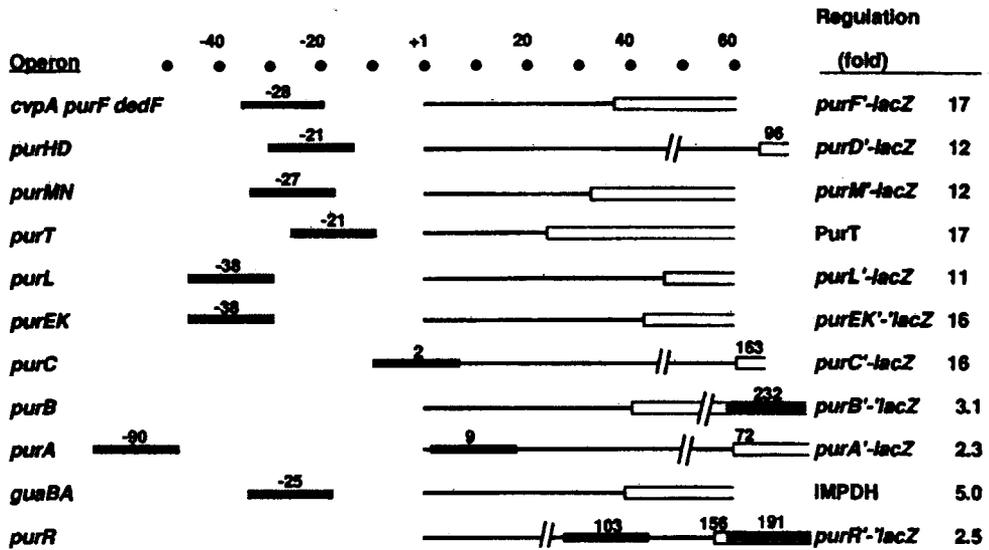


Figure 1.06 The *pur* operon. A) Genes regulated by PurR and their locations in pathways in (clockwise from upper left) *de novo* purine synthesis, *de novo* and salvage pyrimidine synthesis, polyamine synthesis, and self-regulation. B) Alignment and homology among the full-site PurR binding sites. The seventh position in each half-site is starred both above the figure and below the numbering scheme; the best characterized *purF* binding site is at the top, the consensus perfect palindrome used in many crystallography studies to overcome statistical disorder is at the bottom. C) Variety of the ability of PurR to repress different genes (right hand column) is attributed both to sequence variation in the binding site and its location with relation to the start site for transcription; operator name is to the right, numbers over black bars indicate the binding site's distance from the transcriptional start site (open rectangles.) Adapted from Zalkin and Nygaard (1996). (Zalkin 1996)

B

	*            *
<i>purF</i>	ACGCAAACGTTT <b>T</b> CTT
<i>purMN</i>	TCGCAAACGTTT <b>G</b> CTT
<i>purT</i>	ACGCAAACGTTT <b>T</b> CGT
<i>purL</i>	ACGCAAACGTTT <b>T</b> CGT
<i>purEK</i>	ACGCAA <b>C</b> CGTTT <b>T</b> CC <b>T</b>
<i>purC</i>	ACGCAAACG <b>T</b> GT <b>G</b> CGT
<i>purB</i>	ACGCA <b>A</b> T <b>C</b> CGTT <b>A</b> CC <b>T</b>
<i>purHD</i>	GCGCAAACGTTT <b>T</b> CGT
<i>purA1</i>	AG <b>G</b> AAAACG <b>A</b> TT <b>G</b> GC <b>T</b>
<i>purA2</i>	AAGCAAACG <b>T</b> G <b>A</b> TT <b>T</b>
<i>guaBA</i>	ATGCAATCG <b>T</b> T <b>A</b> CG <b>C</b>
<i>purR1</i>	AGGCAAACGTTT <b>C</b> CA <b>C</b>
<i>purR2</i>	GAGCAAACGTTT <b>C</b> CA <b>C</b>
<i>pyrC</i>	AG <b>G</b> AAAACGTTT <b>C</b> CG <b>C</b>
<i>pyrD</i>	CGGAAAACGTTT <b>G</b> CG <b>T</b>
<i>prs</i>	AAGAAAACGTTT <b>T</b> CG <b>C</b>
<i>codBA</i>	ACGAAAACG <b>A</b> TT <b>G</b> GC <b>T</b>
<i>glyA</i>	AG <b>G</b> T <b>A</b> T <b>C</b> CGTT <b>T</b> GCG <b>T</b>
<i>gcv</i>	AAGAGAACG <b>A</b> TT <b>G</b> CG <b>T</b>
<i>glnB</i>	ATGCAAACG <b>A</b> TT <b>T</b> CAA
<i>speA</i>	AAGAAA <b>C</b> CG <b>T</b> T <b>G</b> CG <b>C</b>
perfect	ACGCAAACGTTT <b>G</b> CG <b>T</b>
palindrome	2345678998765432
	*            *

C



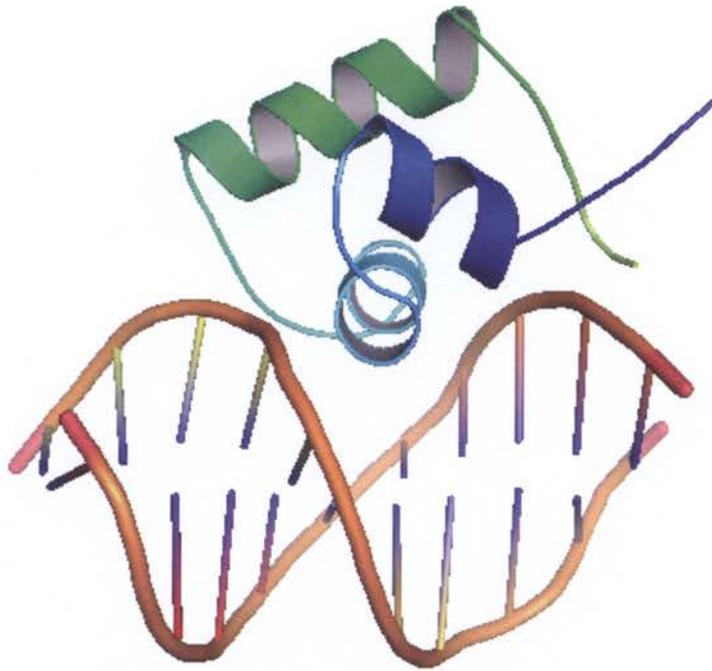


Figure 1.07: The LacI DNA binding domain complexed with DNA (PDB code 1LCD) solved by NMR techniques in 1993 by Chuprina, *et al.* (Chuprina, Rullmann et al. 1993)

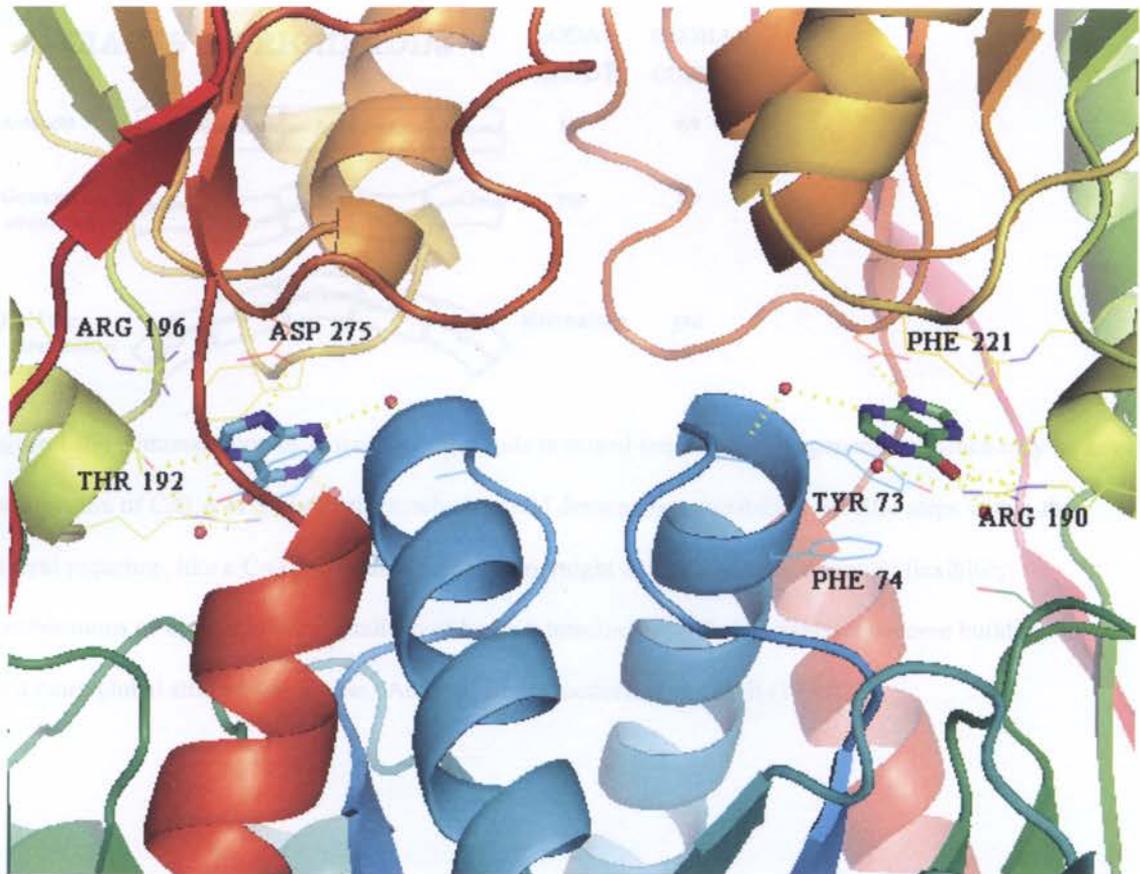


Figure 1.08: hypoxanthine binding pockets at the junction between the N- and C-terminal corepressor binding domains in PurR. Hypoxanthine is shown in ball and stick models, PurR is shown as a cartoon with some ligand binding residues shown in line drawings with labels. Water molecules are red spheres and hydrogen bonds are dashed yellow lines.

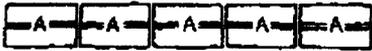
<b>A-TRACTS AS RIGID RODS</b>		<b>LOCAL BEND?</b>	<b>GLOBAL CURVE?</b>
<b>A-tracts</b>		no	no
<b>General sequence</b>		yes	no
<b>Half turn alternation</b>		alternating	yes

Figure 1.09: demonstration of A-tracts as rigid rods in mixed sequences. The general sequence may be any combination of C,G,A or T nucleotides, which would demonstrate flexibility; specific steps within the general sequence, like a C-G pyrimidine-purine step, might evidence a more dramatic flexibility.

Combinations of these structural qualities of bases interacting with their neighbors become building blocks for a more global structure in a gene. Adapted from Goodsell et al., JMB (1994).

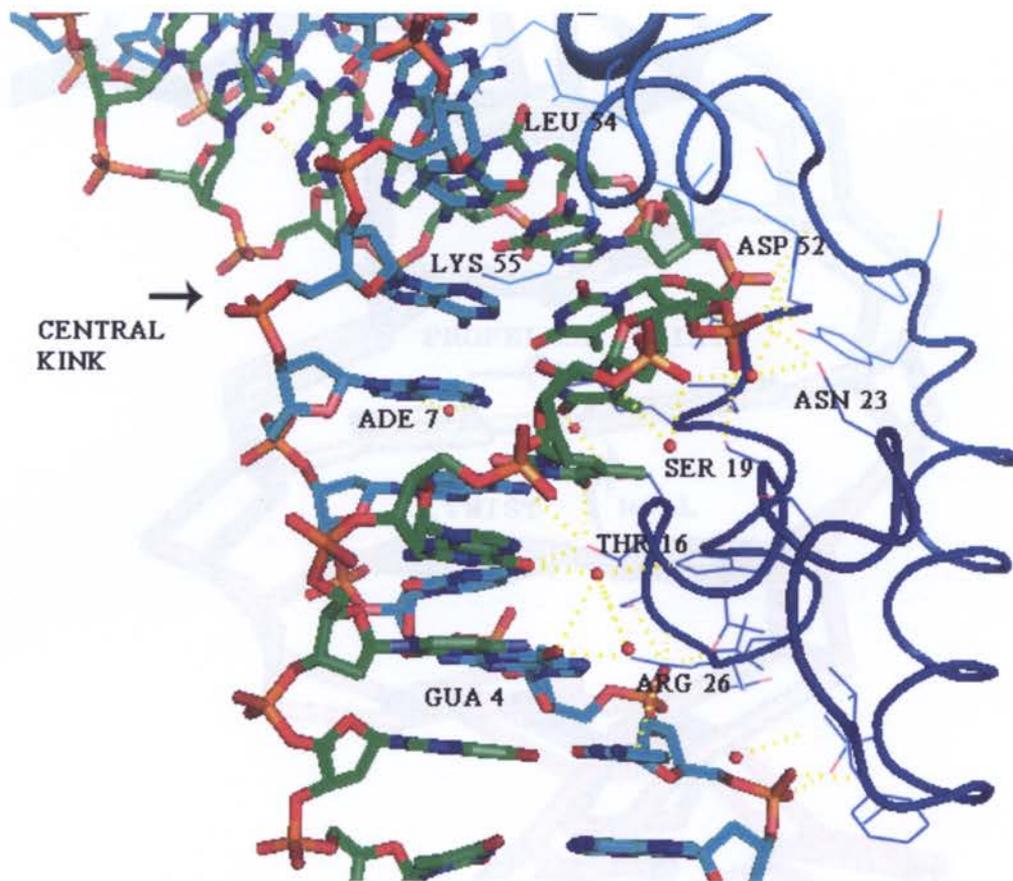


Figure 1.10: PurR-DNA contacts in a single *purF* half-site. The central kink at the CpG step at the edge of the half-site is indicated towards the top of the figure, a single DNA-binding helix-turn-helix domain is illustrated with a narrow cartoon, with residues making hydrogen bonds to the DNA depicted in atom-specific coloring as sticks.

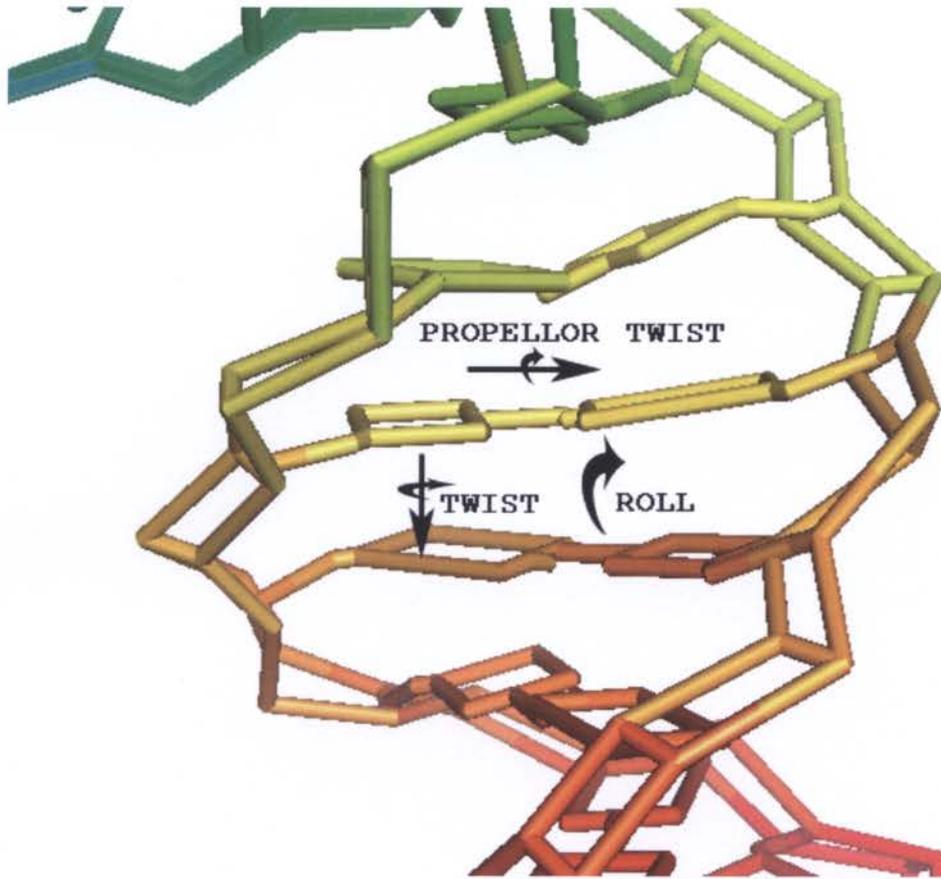


Figure 1.11: cartoon representation of purine and pyrimidine base pairs in the *purF* halfsite (the central CpG step between the half-sites is at the top of the figure, in green). Key terms for the discussion of DNA structure are labeled, including the intra-base pair propellor twist, and the inter-base pair twist and roll. Figure based on the output from the program Curves 5.2.



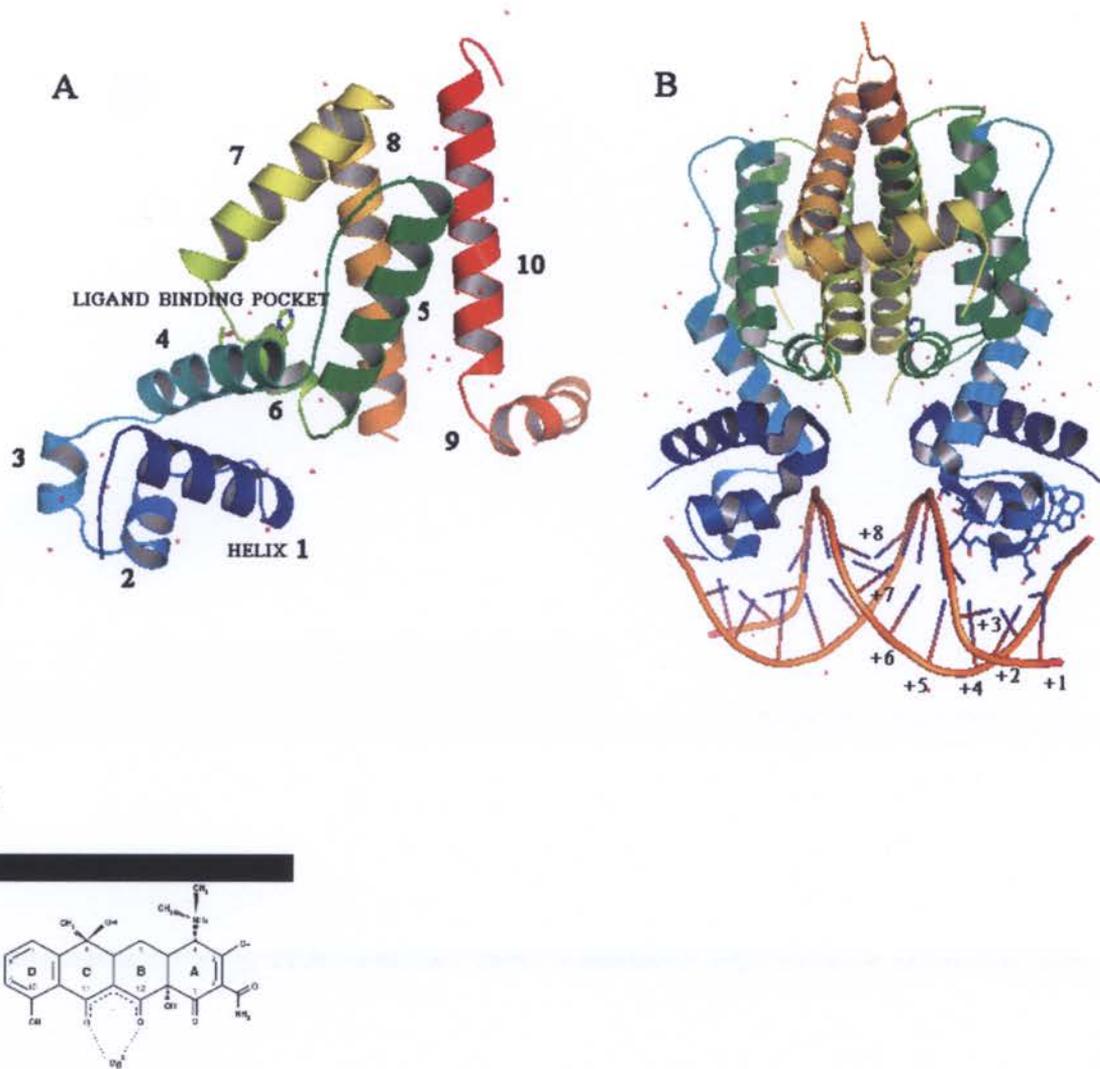


Figure 1.13: Tetracycline and the Tetracycline repressor in apo and DNA bound structures. A) Apo TetR protomer with helices labeled. Residues in the binding pocket that make contact to tetracycline are shown as sticks, there are additional contacts to the loop between helix 4 and 5. B) TetR bound to its 15 base pair operator (PDB code 1QPI). Loops for which there was insufficient electron density to place residues in a specific location are not modeled, instead, the graduated coloring within each monomer indicates where the model picks up the sequence. Base pairs of DNA are labeled, and residues making contacts to the DNA are shown as sticks. C) Tetracycline complexed with Mg ion.

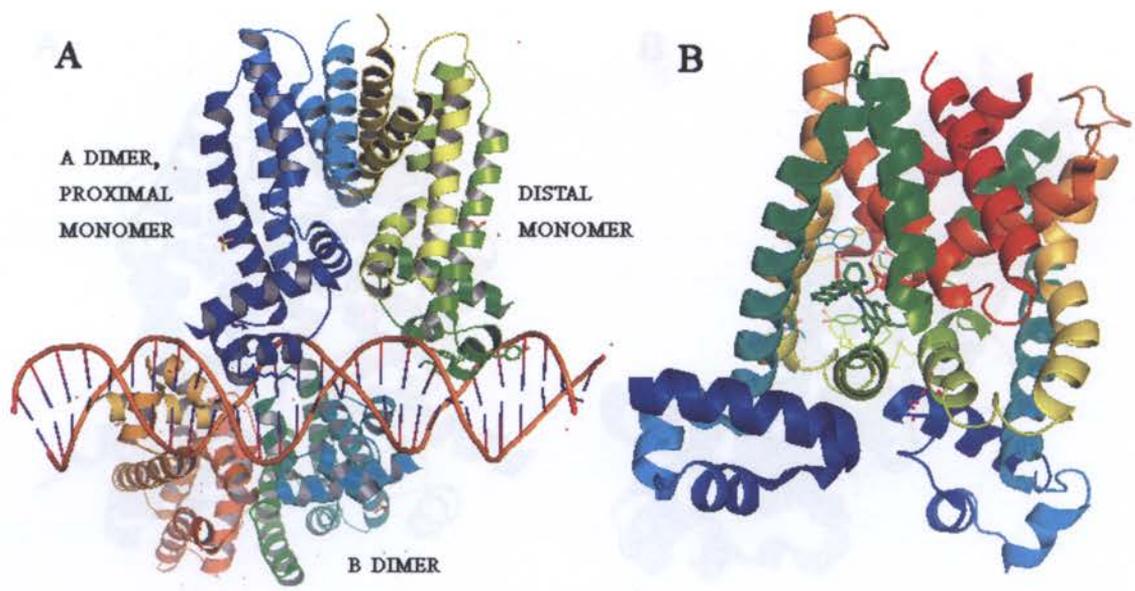


Figure 1.14: QacR structures, bound to DNA and bound to two drugs in the multidrug binding pocket. A) the DNA bound QacR as a dimer of dimers on a 21 base pair operator (PDB code 1JT0). Selected residues in the proximal and distal monomers making contacts to the DNA are shown as sticks. B) QacR bound to two different drugs (ethidium and proflavin) in its remarkable multidrug binding pocket. A series of aromatic residues contributed from both chains of the dimer along the expansive binding pocket allows the protein to bind structurally diverse drugs in a variety of orientations along the pocket. (PDB code 1QVU).

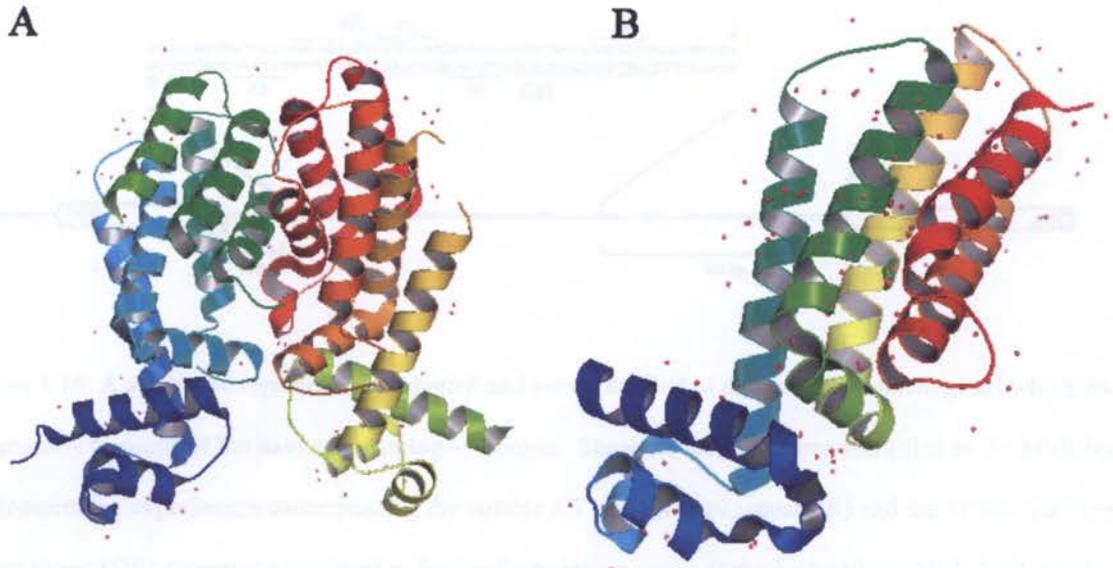


Figure 1.15: The crystal structures of two other TetR family members. A) Homodimer of a gamma-butyrolactone autoregulator receptor protein (CprB) in *Streptomyces coelicolor* A3(2) to 2.4 Å. (PDB code 1UI5). B) Protomer of EthR, a repressor implicated in ethionamide resistance in *Mycobacterium tuberculosis*, to 1.7 Å (PDB code 1T56). Surface plasmon resonance studies indicate a functional DNA binding stoichiometry of 8 protomers/operon.

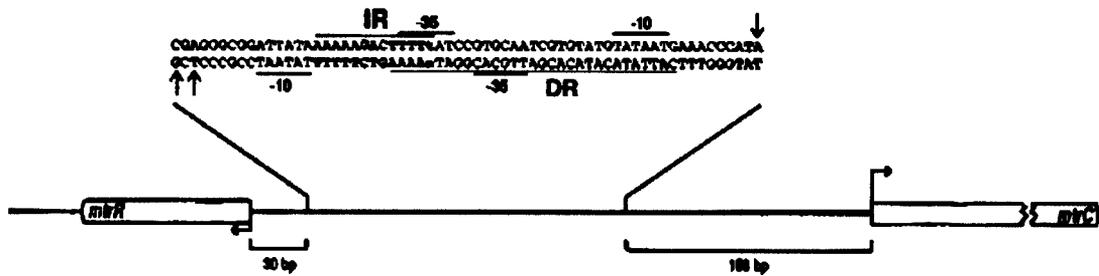


Figure 1.16: A schematic representation of *mtrR* and *mtrC* (the first of the *mtrCDE* tandem gene) which are divergently transcribed but have overlapping -35 boxes. Shown is the broad area identified as the MtrR binding site by footprinting experiments encompassing the notable AT rich inverted repeat (IR) and the 31 base pair imperfect direct repeat (DR) sequence pinpointed in further footprinting assays as the high affinity MtrR binding site. Figure adapted from Hagman and Shafer, *J Bacteriol.* 1995 Jul;177(14):4162-5.

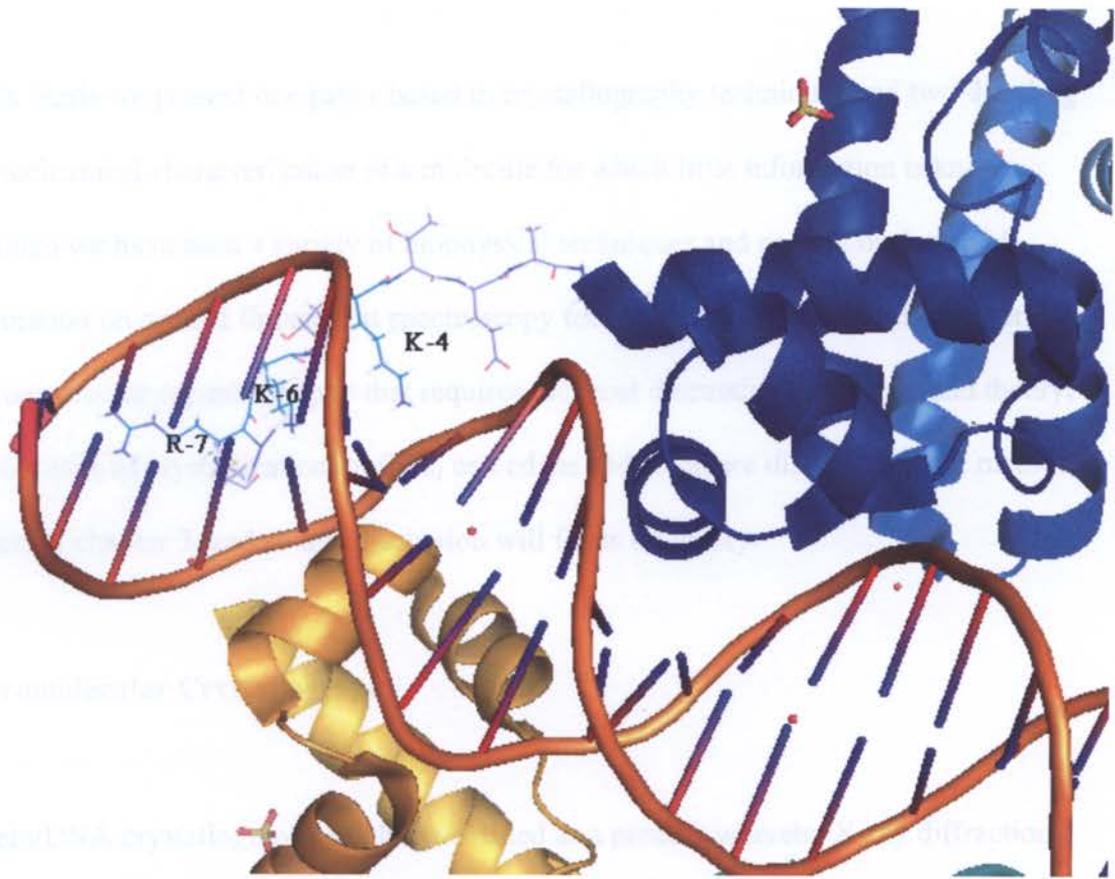


Figure 1.17: a model of the 8 residue N-terminal extension of MtrR on the QacR DNA bound dimer of dimers. The extension is modeled as a random coil making contacts to the minor groove of the DNA, K+4, K+6 and R+7 are highlighted in light blue based atom specific coloring and labeled.

## **Chapter 2: Methods**

In this thesis we present one paper based in crystallography techniques and two detailing the biochemical characterization of a molecule for which little information is known.

Although we have used a variety of biophysical techniques and present background information on several fluorescent spectroscopy techniques at the end of this chapter, it is macromolecular crystallography that requires the most discussion of background theory; practicalities of crystallization, buffers, cell edges and so on are discussed in the methods section of chapter 3, and so this discussion will focus on theory.

### **Macromolecular Crystallography.**

Protein/DNA crystallography might be defined as a process whereby X-ray diffraction patterns, from crystals grown of purified protein and oligonucleotide complexes, are measured and converted to three-dimensional maps of electron density. Into the density maps a model is built or fit, which can then be tested for fidelity to both the electron density and the known structural parameters of protein and DNA and refined. In this process of regular steps, several key points have very interesting considerations, which are discussed in a number of textbooks (Blundell 1976; Stout and Jensen 1989; van Holde, Johnson et al. 1998; Drenth 1999).

The growth of a crystal made up of a pure complex of protein and DNA is a trial and error process of maintaining the fold and activity of the complex while creating an

ordered, three-dimensional array of the complexes, repeating periodically in all directions. The most common approach to creating crystals is to bring a homogeneous macromolecule solution to supersaturation in order to force the molecule out of solution by strategically excluding the molecule from the bulk solvent, disrupting the hydration layer or depriving the molecule of ions or solvent (figure 2.01). This is done by the addition of precipitating reagents like high salt or polyethelene glycols, or by decreasing the dielectric properties of the solution with the addition of organic solvents. Initially, the solution is added to an equal amount of protein solution, but in the presence of a large reservoir of full-strength solution in a closed environment to allow diffusion to increase the pressure on the protein to remove itself from solution. The hope is that the protein will do this by the formation of stable nuclei, and that in the saturated, nucleated solution this produces, the nuclei will grow with further addition of molecules in an ordered way into crystals of at least a 10-20 micrometers in any given dimension.

PurR/hypoxanthine/*purF* crystals were initially grown in a somewhat unusual way: instead of nucleating in a homogeneous solution of protein complex and mother liquor and growing into 3 dimensional crystals, the PurR complex crystallized in groups of plates at the intersection of layers in a biphasic drop. The crystals then melted back into a locally supersaturated pocket and rearranged themselves into the three dimensional crystals used to solve the structure. The process took several months to complete and was not practical for our purposes of obtaining complexes that were not particularly tightly bound together. Instead, we employed a process called seeding, in which a three dimensional crystal of the wild-type complex was manually crushed and serially diluted

into a crystal growing solution containing the mutant complex of interest; these surrogate nuclei allowed us to bypass the plate clusters and grow diffraction quality crystals on a much quicker timeline (a matter of weeks).

Once a crystal of appropriate size and diffraction quality is grown, one must ensure that the crystal contains all the molecules of the complex, is large enough to provide measurable data with reasonable error, and is stable enough to withstand the sometimes long and taxing process of collecting diffraction data. Some of these parameters are out of the control of the crystallographer, but some can be manipulated, like temperature, hydration of the crystal, use of protecting reagents to increase stability, and manipulation of growth conditions are all variables. The single most effective way to improve the odds of a crystal maintaining integrity throughout the data collection process, however, is to ensure that the data collection doesn't go on any longer than necessary. One does this by taking advantage of the crystal's natural symmetry.

#### Crystal symmetry:

At it's most basic, a unit cell is defined by three edge lengths,  $a$ ,  $b$  and  $c$ , and the angles between them  $\alpha$  (between  $b$  and  $c$ ),  $\beta$  (between  $a$  and  $c$ ), and  $\gamma$  (between  $a$  and  $b$ ) as seen in figure 2.02. With these parameters, one can define 7 crystal systems by whether and which parameters are unique and which are equal. The lowest symmetry (no equivalent edges or angles) is the triclinic cell; the rest are monoclinic, orthorhombic, trigonal, tetragonal, hexagonal and cubic, having some or all edges and angles equivalent. The PurR/hypoxanthine/*purF* complex crystals discussed in Chapter 3 grow in an

orthorhombic lattice, where  $\alpha = \beta = \gamma = 90^\circ$  and a, b, and c are independent of each other and must be determined.

Lattices within these crystal systems allow for further definitions of symmetry and therefore further simplification when it comes to practical application. Lattices (called Bravais lattices, figure 2.03) at their most basic are established by points at the corners of the unit cell in the seven primitive (P) cells already mentioned, and additional lattice points can also be defined in non-primitive cells; points definable on opposite faces of the cell in the bc, ac or ab planes are termed A, B and C lattices. If all the faces contain centered lattice points it is the face centered (F) lattice, and a single point centered within the three dimensional unit cell defines an interior (I) lattice. Crystal systems and lattices allow a crystallographer to define the unique information in the crystal and take advantage of symmetry to simplify data collection; to that end, the origin of the unit cell, and by extension the edges, angles and lattice, is defined to take maximal advantage of symmetry elements, like rotary axes, mirrors, centers of symmetry and inversion axes. These symmetry operators are termed point groups, and there are 32, denoting each of the possible unique combinations; combined with the 14 possible Bravais Lattices, there are a finite number of ways (230) to describe an infinite lattice of repeating identical objects as space groups. Protein molecules, however, have intrinsic symmetry of their own, and cannot take the structure of a D-isomer. As a result, mirror planes and inversion axes are not applicable symmetry elements in a protein crystal lattice, and the number of possible space groups which could describe a protein crystal drops to 65: those without symmetry, or with combinations of rotary or screw axes. The PurR/hypoxanthine/*purF* crystals are

described by the  $C222_1$  spacegroup: a center latticed (in the *ab* plane), orthorhombic cell in which there are 2 two-fold axes (one of which coincides with a crystallographic two-fold axis about *x* as well, called a special position, because PurR is a homodimer binding two hypoxanthine molecules symmetrically and a pseudosymmetric DNA oligonucleotide), and one  $2_1$  screw axis. The PurR/hypoxanthine/*purF* crystals therefore have 8 equal molecular assemblies (called asymmetric units) comprising one PurR monomer bound to one molecule of hypoxanthine and one DNA halfsite and related by symmetry to each other in the unit cell.

The molecular assemblies being equivalent in terms of the greater unit cell symmetry, however, doesn't mean that they are identical as biological macromolecules; making changes to only one half-site in the DNA oligonucleotide, as we have done in our studies, leads to a phenomenon called statistical disorder, in which, due to symmetry, the structure solved and refined and analyzed is actually an average of the half-sites. Some analysis (particularly of the more global changes to the structure) can be usefully done, especially as single half-site mutations are biologically and functionally relevant, but ultimately the same mutation on both half-sites will need to be done to minimize error and show detail conclusively at the point of the base change.

Nevertheless, the orthorhombic symmetry and space group of the PurR/hypoxanthine/*purF* crystals mean, in a practical sense, that significantly less data needs to be collected than would be necessary for a triclinic cell. The practical payoff of

knowing how to capture the unique information in a crystal is ironic considering that it is the repetitiveness of the information that makes detection of diffraction possible.

### X-ray Diffraction.

A beam of x-rays produced by bombarding a metal target (in the case of our in house system a copper rotating anode), when encountering a crystal, enters the affair in a focused, monochromatic beam, and leaves it with various waves pointed in different directions, out of phase depending on when and where and what they hit, to be caught and recorded in their new orientation on a piece of film or a detection plate on the other side of the crystal. One might surmise, from the goal of electron density maps, that it is electrons that diffract an X-ray wave off its course; the reason for this is given by the equation for Thompson coherent scattering:

$$\text{Intensity}_{\text{scatter}} = \text{Intensity}_{\text{incident}} * \frac{n e^4}{2 r^2 m^2 c^4} * (1 + \cos^2 2\theta)$$

In which mass (m) is located on the bottom of the equation and therefore has an inverse relationship with the ability to diffract (n is the effective number of independently scattering electrons, e is the charge of an electron, r is the distance from the scatterer, c is the velocity of light and  $\theta$  is the angle of incidence). Protons, by contrast, are about 2000 times heavier than an electron and absorb too much energy for the interaction to be elastic; it is electrons that diffract (Drenth 1999). Because X-ray waves in phase are additive, and an X-ray beam hits an ordered array of electrons in a front, equivalent electrons in different unit cells will each diffract an X-ray wave that maintains its phase, and if there are enough equivalent phased waves at the same angle, they will register on the detection plate as a discrete, measurable, spot. However, not every electron diffracts

an X-ray at any given moment during data collection; the ones that do satisfy the equation:

$$\text{(Bragg's Law)} \quad 2d\sin\theta = n\lambda$$

Where  $d$  is the distance between planes in the crystal lattice,  $\theta$  is the angle of the X-ray beam incident to the lattice plane,  $n$  is an integer and  $\lambda$  is the wavelength of the X-ray.

(Figure 2.04) Bragg's law may be rearranged to the form  $\sin\theta = (n\lambda/2)(1/d)$ , revealing an inverse relationship between  $\sin\theta$  and the interplanal distance of the lattice. To facilitate interpretation of X-ray diffraction patterns, a theoretical construction called a reciprocal lattice is used to simplify the relationship and visualize the direction of scattering more intuitively.

Reciprocal space is defined by  $h, k$  and  $l$ , normal to the real space lattice planes, radiating from an arbitrary origin lattice point, the length of  $1/d_{hkl}$  (the perpendicular distance between sets of  $h, k$  or  $l$  planes, figure 2.05). Reciprocal space allows us to describe Bragg's law of diffraction in terms of an Ewald sphere (figure 2.06), where a sphere of radius  $1/\lambda$ , center of  $M$ , and all points  $P$  equivalent to a reciprocal lattice point of origin  $O$  on the sphere satisfy Bragg's Law, and will diffract the incident beam  $s_0$  at angle  $2\theta$  provided the beam passes through  $M$  and  $O$ . From this construction, it is possible to better visualize the number of points in reciprocal space not satisfying Bragg's Law with all parties in fixed position; however, should the crystal (and therefore the reciprocal lattice) be slowly rotated through the Ewald Sphere, each point would at some point satisfy Bragg's Law, and one would need only collect enough degrees of rotation sufficient to represent the unique data in the unit cell. While diffraction, beam angle and

position is enough to calculate the dimensions of the unit cell, however, it is the intensity of the diffraction that is useful for solving for the structures within the unit cell, which is ultimately the information of interest.

### The Structure Factor Fourier Series.

Because X-rays are periodic and additive, each diffracted x-ray can be described mathematically by a Fourier series called a structure factor equation. Specifically, the structure factor ( $F_{hkl}$ ) is the sum of a series of terms describing the contribution of each atom to the overall reflection.

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

Where  $f_j$  is the scattering factor of the  $j$ th atom,  $hkl$  and  $x_j$ ,  $y_j$  and  $z_j$  are the reciprocal and real space indexes for that atom, respectively. In this equation, an electron is assumed to be a simple sphere of density, and  $F_{hkl}$  can also be described in terms of the electron density ( $\rho$ ).

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

Or in terms of the unit cell's electron density (where  $V$  = volume of the unit cell),

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

Structure factor amplitudes ( $|F_{hkl}|$ ) are measurable, since they are the square of the intensity; structure factor phase information ( $\alpha_{hkl}$ ) is lost in the recording of the wave on a two-dimensional detection device. Ignoring the phases for a moment, we can use the

reversibility of Fourier series to convert the equation to solve for electron density, which is the function of interest.

$$\rho(x,y,z) = (1/V) \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i(hx_j + ky_j + lz_j) + i\alpha_{hkl}}$$

And at this point, the only missing information is the phases, which must be borrowed, copied or estimated.

### The Phase Problem.

Techniques to determine the phases of a structure depend on whether there is useful information already available on the structure within the unit cell. In the structures discussed in Chapter 3, there was a great deal of information already available on the structures since we were exploring the effects of point mutations, and as a result we used a technique called molecular modification, in which the phases from the original structure were lifted and then refined to the new structures. This is by far the most straightforward method of solving the phase problem, but as previous information does not always exist, we shall briefly describe techniques for solving the phases using an educated guess and using no previous information at all.

To solve the phases in the absence of any starting information about the structure within the cell, one of two methods might be employed, based on a similar theory. To visualize this crystallographers often rely on an Argand diagram depicting the real component of the structure factor on x, or  $F(\cos\alpha)$  and the imaginary component on y, or  $F(\sin\alpha)$ , to describe the vector of the structure factor (F) with the length of the vector being the

amplitudes and the angle  $\alpha$  the phases. In the absence of phase information, the potential vector end point might lie anywhere along a circle inscribed about the origin and with the radius of the structure factor amplitudes (figure 2.07). Multiple Isomorphous Replacement (MIR) techniques to solve *de novo* phases take advantage of the similar electron number of carbon, nitrogen and oxygen atoms (six, seven and eight electrons, respectively) to deliberately place atoms of higher electron shell, like metals, in fixed positions in the same unit cell and map them (using difference Patterson functions) as beacon starting points for refining the wild type phases.

The technique was described by Harker as a construct of two overlapping inscribed Argand diagram circles shown in figure 2.07 where, in the heavy atom instance, the origin is slightly displaced off of (0,0) by the structure factor of the heavy atom, and the two circles overlap at only two points; if the crystal, unit cell and space group are isomorphous to the wild-type, then the structures are nearly identical and the vectors must terminate at the same point. The possibilities for the phases of the wild-type vector are therefore narrowed down to two; a second derivative will establish the more likely starting point, and a third heavy atom derivative (if it can be found) might help to overcome error inherent in the experiment. Drawbacks to this technique include the need for nearly identical unit cell dimensions (heavy atom solutions are often soaked into wild-type crystals to try to maintain the cell), the need to fix the heavy atom into position (the ability to incorporate bromine and iodine into nucleotides is often utilized to guarantee position with covalent bonds; alternatively metals like mercury are known to coordinate

well with cysteine residues, and sometimes cysteines are mutated into a structure to encourage this) and the inherent fragility of crystals not standing up to manhandling well.

Multiple Anomalous Diffraction (MAD) is a similar technique that takes advantage of tunable wavelengths from high end X-ray sources and the anomalous scattering abilities of atoms like selenium, to the same effect as heavy atom derivatives in MIR: the anomalous scattering offsets the origin of the vector far enough that the circle describing every possible combination of phases with known amplitudes only intersects with the wild type at two points. Data collection at three wavelengths serves the same purpose as multiple heavy atom derivativitized crystals in MIR, with the advantage of being able to collect them all on the same crystal, and thereby avoiding the difficulty attaining isomorphous unit cells. The requirement that the crystal hold up to three data collections rather than only one is the tradeoff.

A greatly simpler technique is to identify another molecule of very similar secondary and tertiary structure (this can sometimes be done with single domains) and use it as a model to fit into the experimental amplitudes and forward and back calculate phases until eventually a set is found that can serve as starting phases for the experimental structure. This technique, called Molecular Replacement, requires a series of random orientations and locations of the known structure around the experimental unit cell, and a corresponding correlation coefficient (related to the structure factor) to help place the molecule in the cell. One program designed to run such calculations is EPMR, or the Evolutionary Program for Molecular Replacement, for its evolutionary algorithm

(Kissinger, Gehlhaar et al. 1999). The algorithm works by optimizing all three rotational and three translational parameters at the same time, taking the highest scoring correlation coefficients and using the positions that generated them as a starting point for the next round of less random orientations and positions to test. This technique is particularly useful if models of other proteins in the same family exist, and if the global fold of the protein is conserved within the family, even small adjustments like the angle of a helix relative to another or high random coil content can sabotage molecular replacement, however. Eventually, all of these techniques serve to provide a starting set of phases that can be used to visualize the electron density in maps and build a model into it.

Difference Fourier electron density maps compare the structure factors observed from the experimental data ( $F_{obs}$ ) and those back-calculated from the model built into the electron density ( $F_{calc}$ ); two levels of maps are commonly used, one is an  $f_0 - f_c$  map, which is a direct comparison and helpful in the early stages of refinement when large adjustments are still being made, and the other is a  $2f_0 - f_c$  map, which gives weight to the observed structure factors and reflects less model bias.

$$2f_0 - f_c \text{ map: } \rho(x,y,z) = (1/V) \sum_h \sum_k \sum_l \left| 2F_{obs} - F_{calc} \right| e^{-2\pi i(hx_j + ky_j + lz_j) + i\alpha_{hkl}}$$

$$f_0 - f_c \text{ map: } \rho(x,y,z) = (1/V) \sum_h \sum_k \sum_l \left| F_{obs} - F_{calc} \right| e^{-2\pi i(hx_j + ky_j + lz_j) + i\alpha_{hkl}}$$

Electron density maps and difference maps are key to judging the refinement process to judge fidelity to experimental data.

## Refinement.

Refinement begins with the model as a rigid body, eventually to progress to several rigid groups in sections (perhaps by domain) and then eventually to refinement of individual residues, atoms, coordinated solvent molecules and any other ions that may have found a coordinated position near the model fixed enough to diffract electrons. Refinement must take into account the basic stereochemical restraints of protein and nucleotide structure, as well as establishing, via least-squares computations, the location of the atoms that will most satisfactorily account for the electron density observed.

Early refinement programs like the TNT suite of programs (Tronrud 1997) used Fast Fourier Transforms to minimize functions using least-squares refinement. Later software packages like the widely used Crystallographic and NMR System for structure determination (CNS, (Brunger, Adams et al. 1998)) use a technique called simulated annealing in combination with energy minimization to relax the restraints on the structure by simulating high temperatures and assuming higher flexibility and movement; as the protein is allowed to anneal in the simulation, the model overcomes local energy minima to situate more effectively in the electron density at biological temperatures. Energy minimization considers non-covalent bonds such as hydrogen bonding, and brings them into low-energy states within the observed data.

It is useful to begin with moderate resolution data cutoffs (3.5 – 4.0 Å) to start with a larger target in initial rounds of refinement and lock the structure into it with higher resolution detail only after one is sure of the more global structure. Lower resolution

allows for refinement to begin with simple positional x, y, z refinement and add more complex parameters later. With the inclusion of 2.0 – 3.0 Å data can one consider the thermal factor (B-factor) to account for the motion of each atom independently. The B-factors for a string of atoms can give a sense of whether a loop or residue is particularly flexible, for example, or whether the observed position is tightly held: sometimes an indication that an atom modeled as a water molecule might be an atom of higher electron shell if the density for it is high enough to drop the B-factor considerably. A thermal factor of 30-40 Å<sup>2</sup> is a reasonable number, resulting from both the breathing motion of biological macromolecules, variations in structure from molecule to molecule within the crystal, and variations in position within the unit cell as well. The measure of the motion of an atom (the thermal factor) is given by:

$$B = 8\pi^2 \langle \mu^2 \rangle$$

Where  $\langle \mu^2 \rangle$  is the mean squared displacement of the atomic vibration. B factors are often directly related to how well a particular area of the structure diffracts because increased motion will sabotage the phasing of the X-rays adding together in diffraction, and so B-factors are often a good measure of whether a particular part of the model is supported by good electron density.

Between rounds of refinement and model fitting into electron density maps, the question becomes when a model sufficiently represents the data, and how one gauges the fidelity of the model in a quantitative way. The numerical measure (in percent) of how well a model matches the experimental data is called the R factor, defined by:

$$R = \frac{\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|} \times 100$$

$$\sum_{hkl} |F_{\text{obs}}|$$

And is a measure of the difference between the calculated structure factors and observed structure factors, relative to the observed. An initial refinement in rigid body might produce an R-factor of somewhere near 50%, with a number higher than that indicating that the model structure is inappropriate or randomly fitting the data. A final R-factor in a good quality model should be below 25%, and in practice is rarely below 15%, owing to errors in data completeness, processing, and various other points along the way. A 15% deviation from the observed data is a very good model.

A variation on the R-factor used in the CNS suite of programs is  $R_{\text{free}}$  and  $R_{\text{work}}$ , where  $R_{\text{free}}$  represents a less biased comparison, since a small sub-set of data (5-10%, usually) are set aside early in the refinement process and not used to refine the model as well as gauge the model's accuracy; so the data is slightly more objective and the number is likely to be a little bit higher than  $R_{\text{work}}$ . Still  $R_{\text{free}}$  and  $R_{\text{work}}$  should be within a few degrees of each other unless there has been significant model bias of the data.

R-factors are frequently reported continuously during refinement, B-factors can point to areas of flexibility or movement, and difference maps assist with manual repositioning of residues and model building, but there is another map function of great assistance in gauging a near-final model's fit to the electron density, and it is called an omit map. Omit maps are generated when a section of the model (5-10%) is deliberately deleted from map calculation to generate a space in which there is minimal model bias to the electron density. Serial deletion of areas and then the composite collection of minimally

biased sections of density to overlay on the model for comparison check are called composite omit maps and done through one of the programs in the CNS suite.

Still, the most common check on the stereochemistry of a protein model is the Ramachandran plot (figure 2.08); a graphical representation of the  $\phi$  and  $\psi$  angles commonly found in the protein backbone based on residue (glycine, for example, has a great deal more flexibility than other residues) and secondary structure (more restrictive than the random coil possible conformations.) A Ramachandran plot can highlight problems in the model, and also can indicate residues participating in unusual contacts for a specific purpose: PurR, for example, has two residues found in distinctly disallowed regions (Asp 275 and Ser 124) because they are involved in hypoxanthine binding in the ligand pocket. All the PurR structures reported in chapter 3 are consistent in the positions of these two key ligand binding residues.

All of these analysis techniques combine to indicate quantitatively when a model has become a good representation of the data, within stereochemical expectations and with minimal model bias. The process of refining a model might be treated as an ongoing process which always has room to improve, but in practice, once a model's R-factors plateau and will not drop further, when there is no density in the map insufficiently accounted for and no further solvent atoms to place in the structure, then the model is finished and the analysis of what one might learn from it can begin.

## **Fluorescence spectroscopy**

For a structural biologist, and X-ray crystallography model might reasonably be considered to be the goal of a project if one wished to have atomic level information on the interactions, relationships and mechanisms of a macromolecule. Failing a three dimensional crystallographic model, or perhaps complimenting it, there are numerous other techniques for exploring structure and function together; some of the complimentary techniques used in the papers presented here other than crystallography are fluorescence spectroscopy based, and discussed in a number of textbooks, among them Principles of Fluorescence Spectroscopy(Lakowicz 1999) and Principles of Physical Biochemistry (van Holde, Johnson et al. 1998).

### Fluorescence anisotropy

Fluorescence is defined as the emission of light from any substance occurring from a singlet excited state, where oppositely spinning paired electrons allow the return of an excited state to ground state without a change in spin, by photon emission. The emission wavelength is offset from the excitation wavelength, and fluorophores are photoselective for polarized light with electric vectors aligned parallel to the transitional moment and molecular axis. This principle, that fluorophores are photoselective and that the extent to which a fluorophore rotates during the excited-state lifetime determines its emitted polarization, is the basis the technique of fluorescence polarization, or fluorescence anisotropy (FA). In an isotropic solution, a fluorophore (we use fluorescein, figure 2.09a) covalently bound to a relatively small molecule, such as a small oligonucleotide

representing a protein binding site, would be oriented randomly and there would be no preferential emission of polarized light as a result. In a solution where protein is titrated in to bind the fluorophore bound DNA, the molecules would be spinning slower and result in a partially oriented population of fluorophores, which can be measured if parallel ( $I_{\parallel}$ ) and perpendicular light ( $I_{\perp}$ ) emission intensities are compared (figure 2.09b). The equations for polarization (P) or anisotropy (r) are given by:

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}$$

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}$$

We have used this technique both in titration experiments to study affinity and steady state experiments to explore stoichiometry; FA has become a powerful functional assay for DNA binding proteins both as a stand alone experiment and comparatively (Lundblad, Laurance et al. 1996). FA fails at binding constants less than 1 nM, usually due to an inability to get substantial fluorescent signal when the target labeled DNA needs to be 1/10<sup>th</sup> of the binding constant for titration, or at constants more than several micromolar, because titrations begin to substantially change the experimental volume and add error. Difficulties can also arise fitting binding curves that exhibit cooperativity or multi-phasic binding.

### Circular Dichroism

While many fluorophores are aromatic, however, any molecule with asymmetry can influence polarized light. A circularly polarized wave with right and left polarized light,

for example, when it comes into contact with a properly oriented right-handed helix, might find the electron path along the helix, dipole and magnetic moment, acting in concert with the ellipticity to the right, and out of concert with the oppositely polarized light (figure 2.10). This difference creates a characteristic emission spectra for an alpha-helix, as it would with other secondary structural elements in a protein or DNA as well. These spectra are used in circular dichroism (CD) to break down a mixed character protein spectra and attempt to predict it's secondary structural elements by percentage. The process is difficult due to the overlapping wavelengths characteristic of secondary structures (figure 2.11) as well as the problem of not being able to easily take a baseline spectra for structural elements like  $\beta$ -turns; nevertheless, as a comparative technique significant information may be gleaned.

### Dynamic Light Scattering

The last technique to introduce is complimentary to the X-ray scattering discussion. Dynamic Light Scattering (DLS) makes use of Raleigh scattering, or incoherent scattering, to measure macromolecules in solution rather than coherent scattering through a crystal. However, in the case of DLS, one measures the variations in light scattering by a macromolecule with time, which fluctuates due to the Brownian motion of the macromolecule. From these measurements the translational diffusion coefficient ( $D_T$ ) of the macromolecule can be calculated, which in turn allows the determination of the hydrodynamic radius of the average scattering particle ( $R_H$ ) via the equation,

$$D_T = kT/6\pi\eta R_H,$$

where  $k$  is the Boltzmann constant,  $T$  is the temperature in Kelvin and  $\eta$  is the solvent viscosity (Garcia de la Torre and Bloomfield 1981).  $R_H$  can be used to estimate the molecular weight by assuming a globular shape and using a number of known samples as markers. In practice this technique has a great deal of error, arising from variations in oligomeric state, bubbles or dust in the solution, thermal variations or possibly aggregates, and we have used it primarily as a way to confirm information gleaned from other experimental techniques as well.

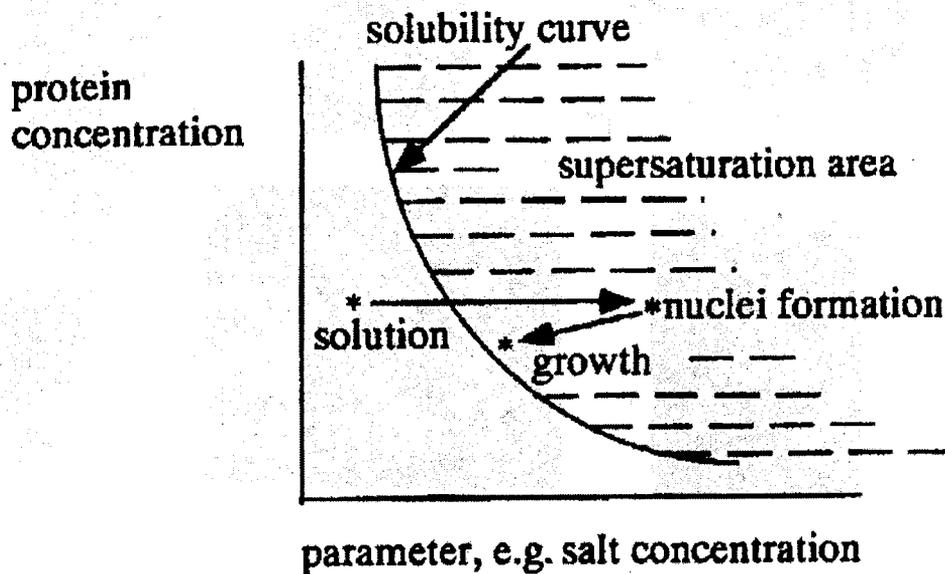
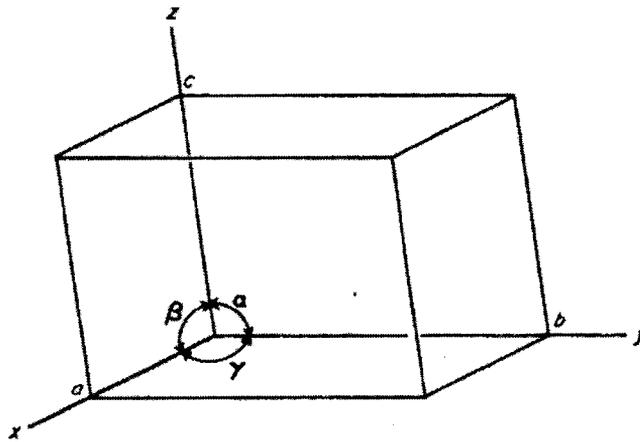


Figure 2.01. A saturation diagram for macromolecule crystal growth. A more supersaturated solution (influenced by precipitants, salt concentration, organic solvents in the mother liquor) is more likely to create nuclei of crystals. If this creates a localized reduction in effective protein concentration, more macromolecules might order themselves onto the nuclei and increase the size of an individual crystal rather than creating a new crystal nucleus. (Drenth 1999)



Crystal System	Number of Independent Parameters	Parameters	Lattice Symmetry
Triclinic	6	$a \neq b \neq c; \alpha \neq \beta \neq \gamma$	$\bar{1}$
Monoclinic	4	$a \neq b \neq c; \alpha = \gamma = 90^\circ; \beta > 90^\circ$	$2/m$
Orthorhombic	3	$a \neq b \neq c; \alpha = \beta = \gamma = 90^\circ$	$mmm$
Tetragonal	2	$a = b \neq c; \alpha = \beta = \gamma = 90^\circ$	$4/mmm$
Trigonal			
rhombohedral lattice	2	$a = b = c; \alpha = \beta = \gamma \neq 90^\circ$	$\bar{3}m$
hexagonal lattice	2	$a = b \neq c; \alpha = \beta = 90^\circ; \gamma = 120^\circ$	$6/mmm$
Hexagonal	2	$a = b \neq c; \alpha = \beta = 90^\circ; \gamma = 120^\circ$	$6/mmm$
Cubic	1	$a = b = c; \alpha = \beta = \gamma = 90^\circ$	$m\bar{3}m$

Figure 2.02: Definitions of a unit cell and the parameters and symmetry that defines cell types. Adapted from Stout and Jensen, 1989 (George H. Stout 1989).

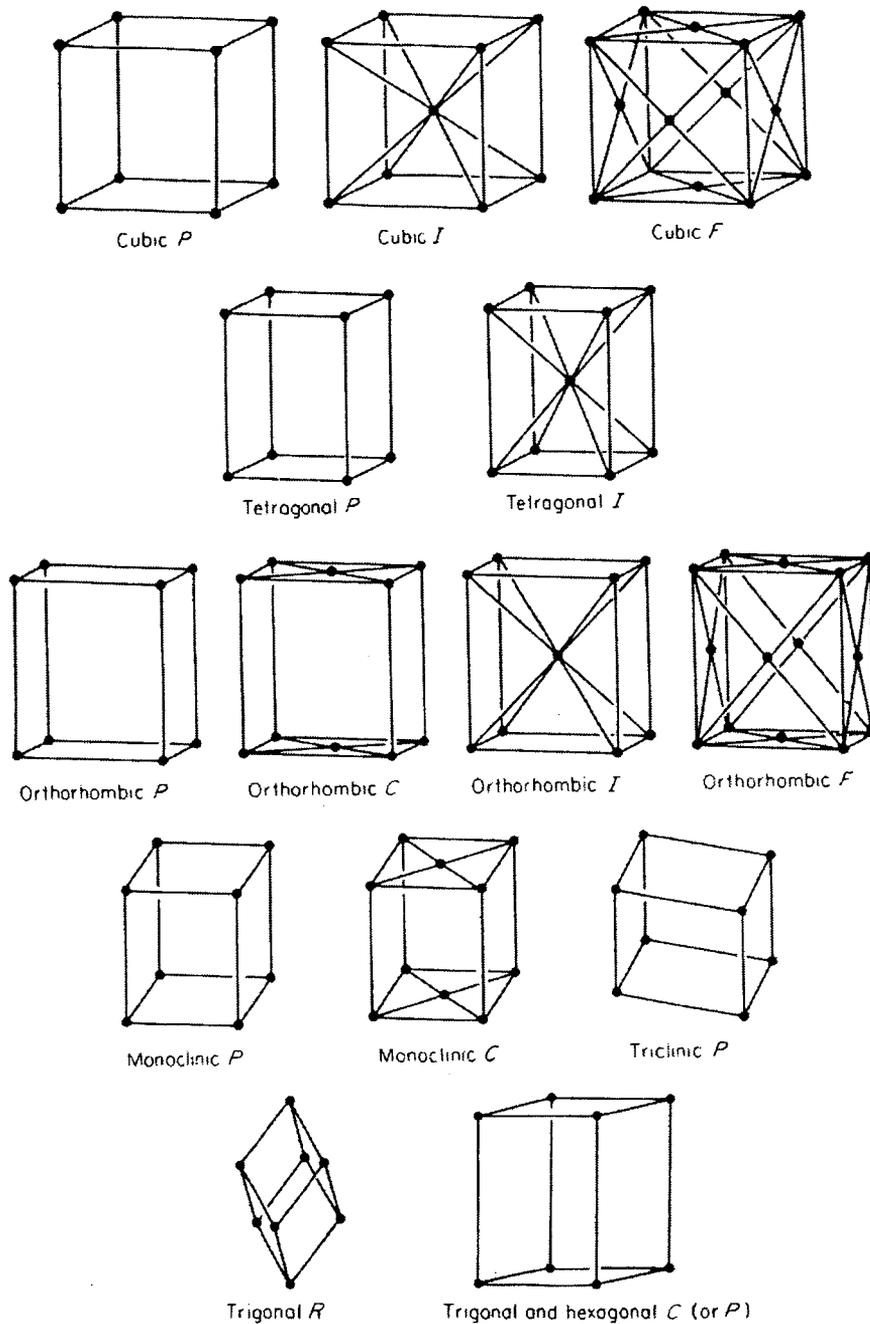


Figure 2.03: The fourteen Bravais lattices, adapted from Blundell and Johnson, 1976 (Blundell 1976).

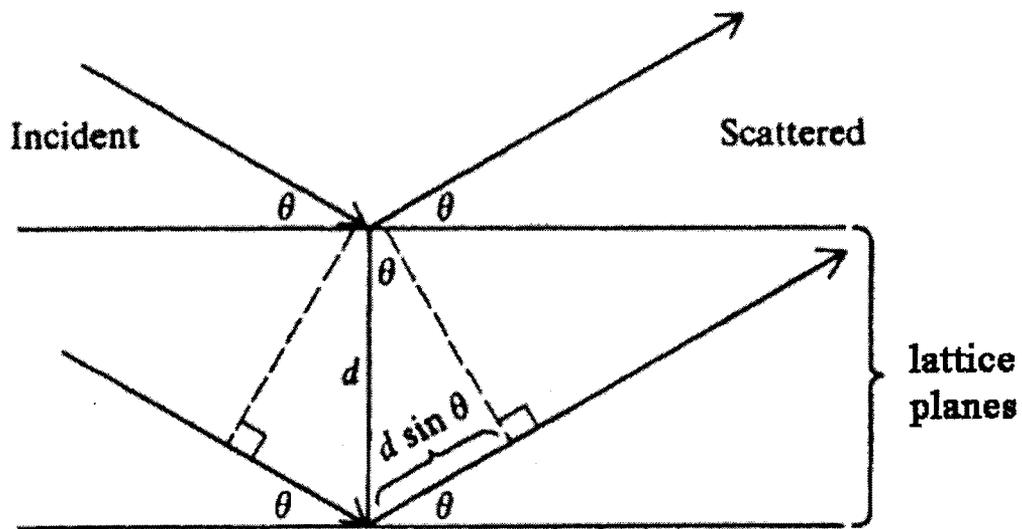


Figure 2.04: An illustration of Bragg's Law, adapted from Drenth, 1999 (Drenth 1999).

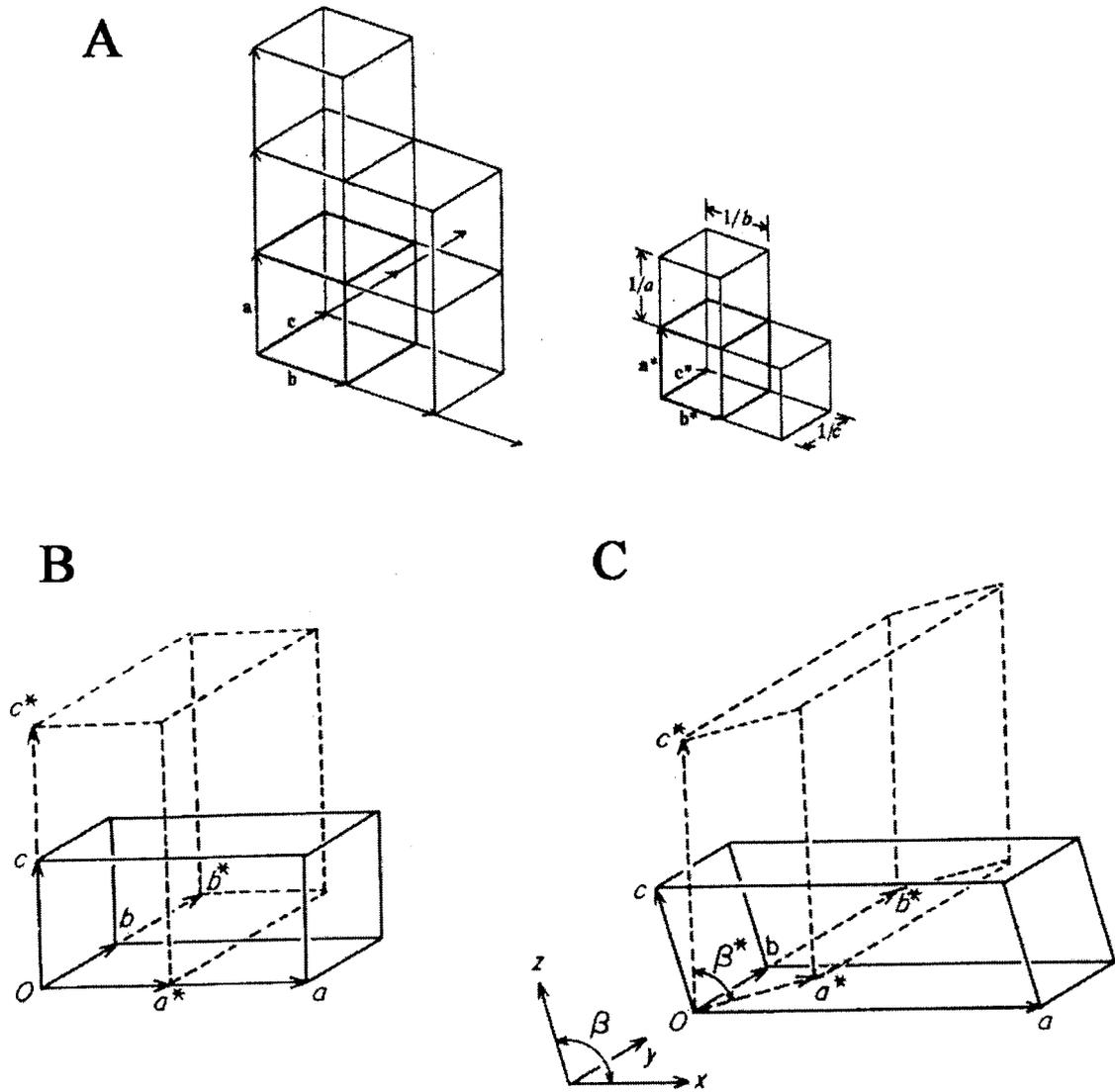


Figure 2.05: Real vs. reciprocal space. A) A cubic unit cell in a lattice where  $a = b = c$  and  $\alpha = \beta = \gamma = 90^\circ$  in real space (on the left) B) An orthogonal cell in which  $a > b = c$ ; real space is in solid lines and reciprocal space is depicted in dashed lines. C) A monoclinic cell where  $\beta > 90^\circ$  and  $a > b = c$ , real and reciprocal space are in solid and dashed lines, as in B). A is adapted from Cantor and Shimmel, 1980 (Cantor 1980), B and C from Stout and Jensen, 1989 (Stout and Jensen 1989).

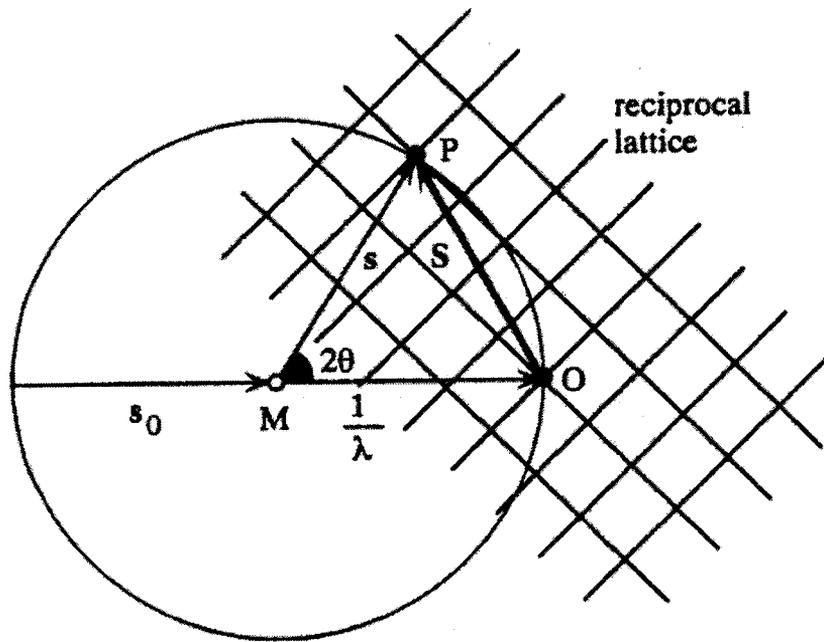


Figure 2.06. The Ewald sphere, with radius  $1/\lambda$ . If the incident beam  $s_0$  passes through both the center of the Ewald sphere (M) and the origin of reciprocal space (O) any other point on the sphere coincident with a reciprocal lattice point P satisfies Bragg's law and will diffract the beam (s). Adapted from Drenth, 1995 (Drenth 1999).

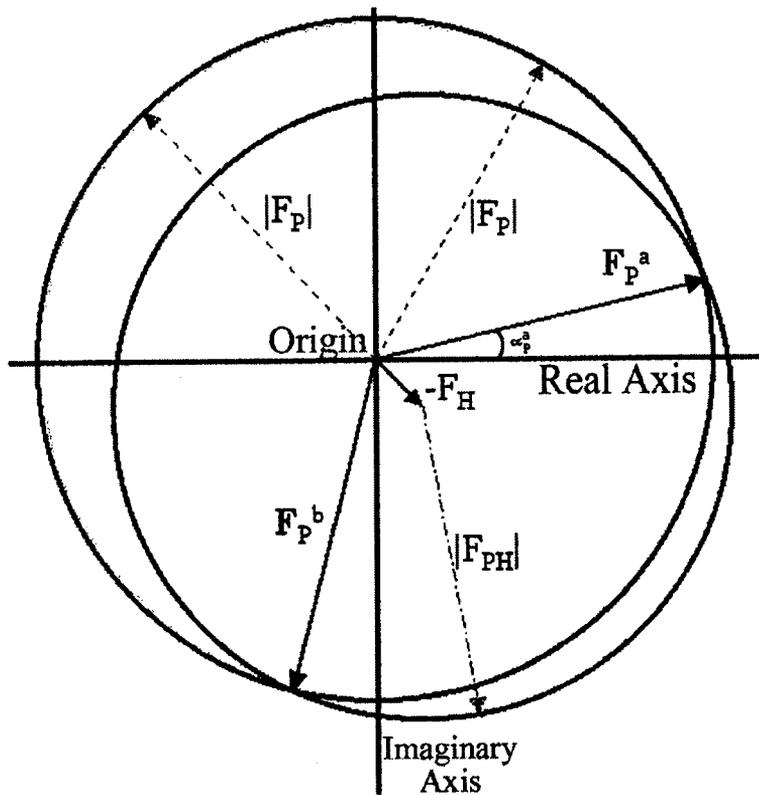
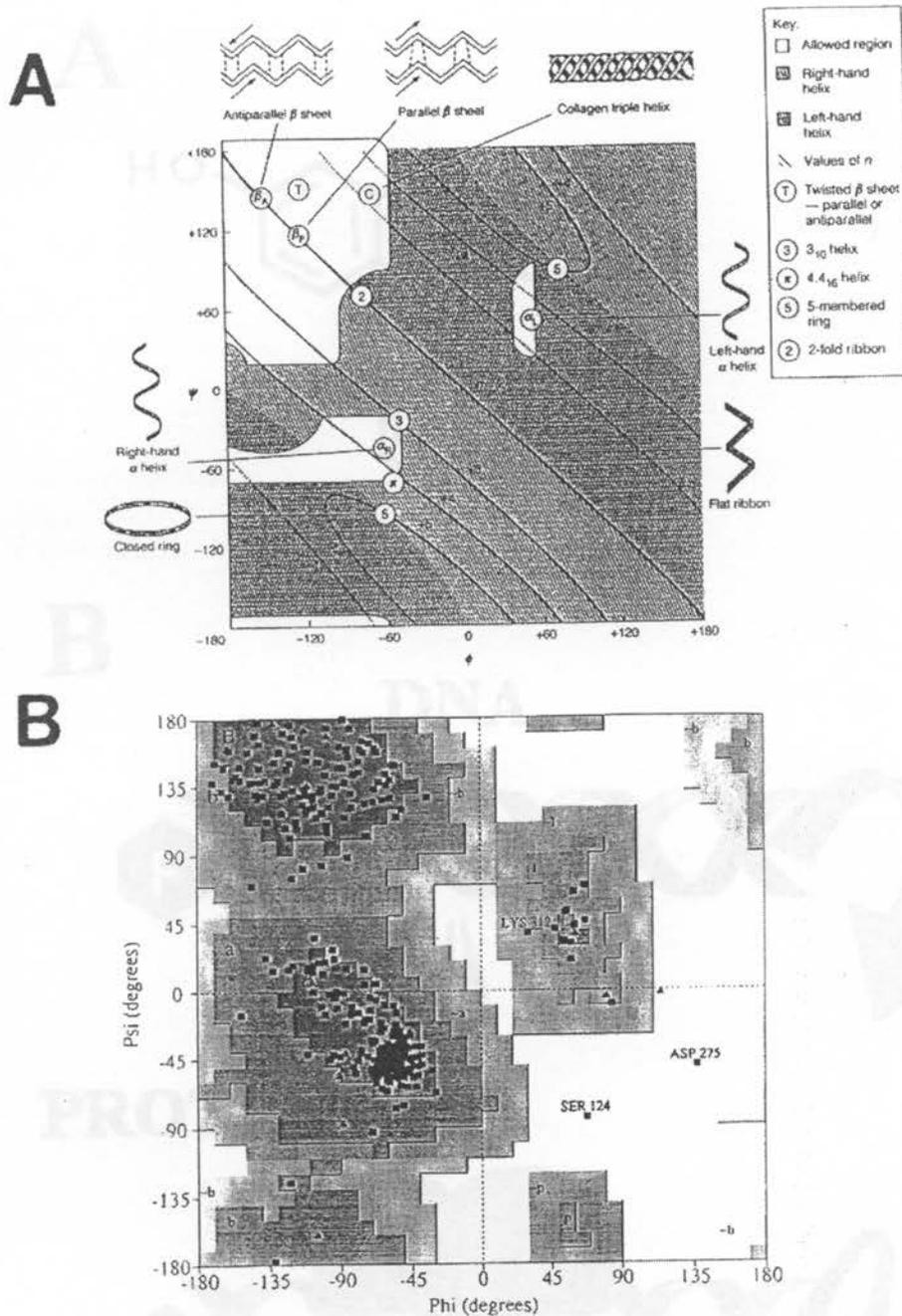


Figure 2.07. Phase determination by the Harker construction.  $F_P^b$  and  $F_P^a$  are the two possible correct structure factor solutions for the phases from this construction.



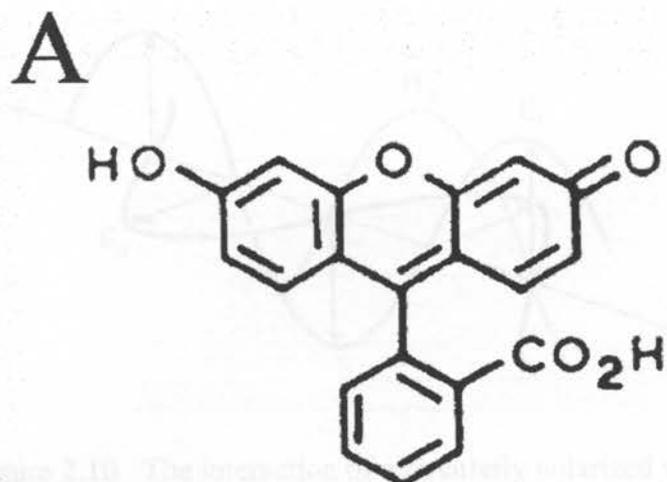


Figure 2.10. The interaction of a linearly polarized wave E(t) in the magnetic field  $H_0$  with a helical molecule would induce the magnetic (m) and electric (e) dipole moments to rotate about  $H_0$ , with the electrons moving in a path indicated by arrows, along its course with one half of the wavelength out of phase with the other.

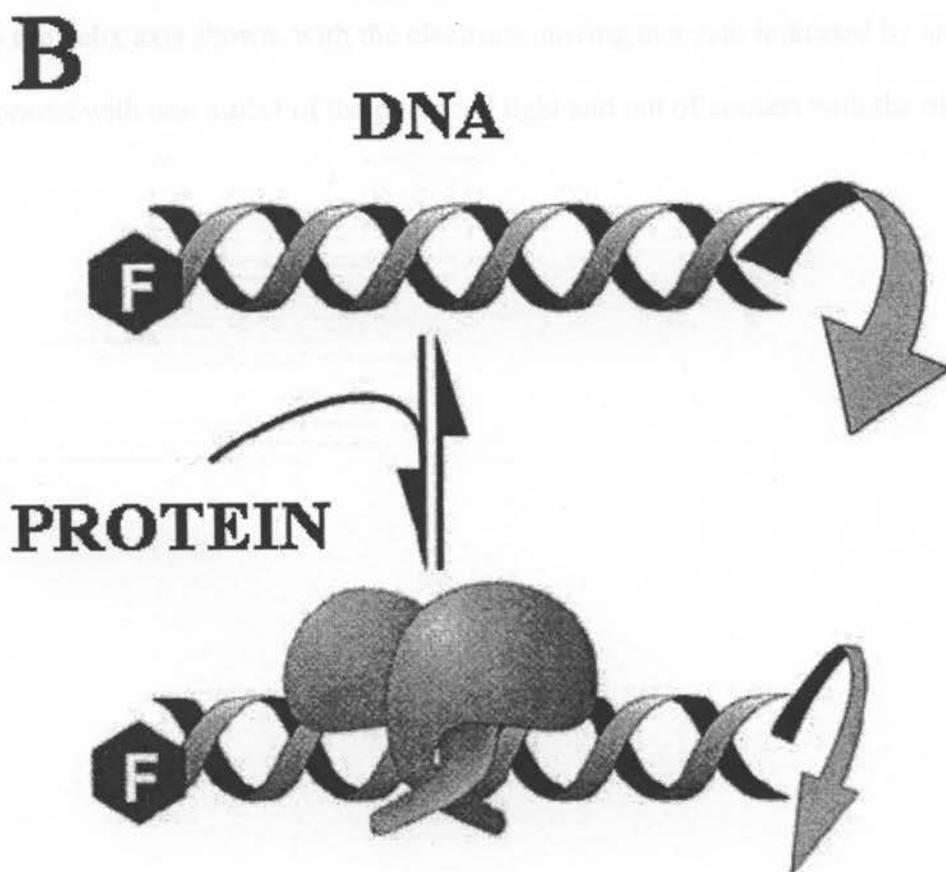


Figure 2.09. The chemical structure of fluorescein, and the spin associated with binding in an anisotropy experiment.

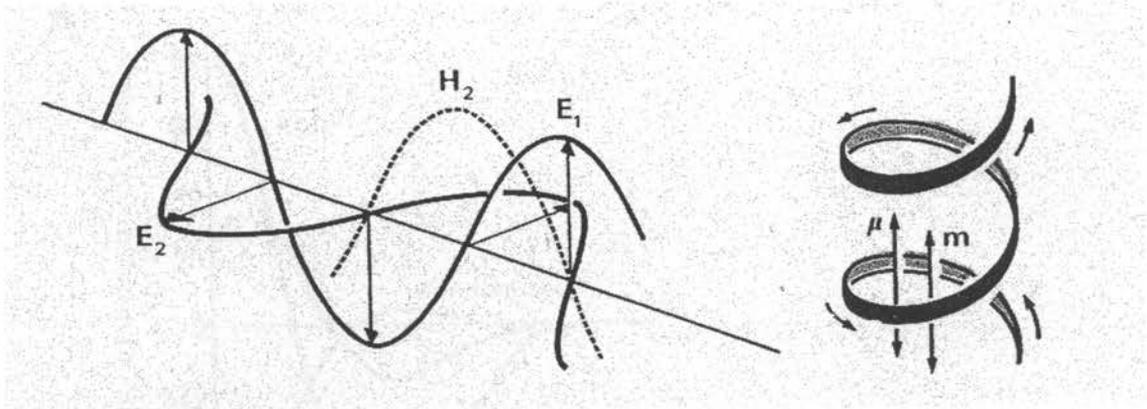


Figure 2.10. The interaction of a circularly polarized wave  $E$  ( $H_2$  is the magnetic field) with a helical molecule would induce the magnetic ( $\mathbf{m}$ ) and electric ( $\mu$ ) moments parallel to the helix axis shown, with the electrons moving in a path indicated by arrows, acting in concert with one aspect of the polarized light and out of concert with the other.

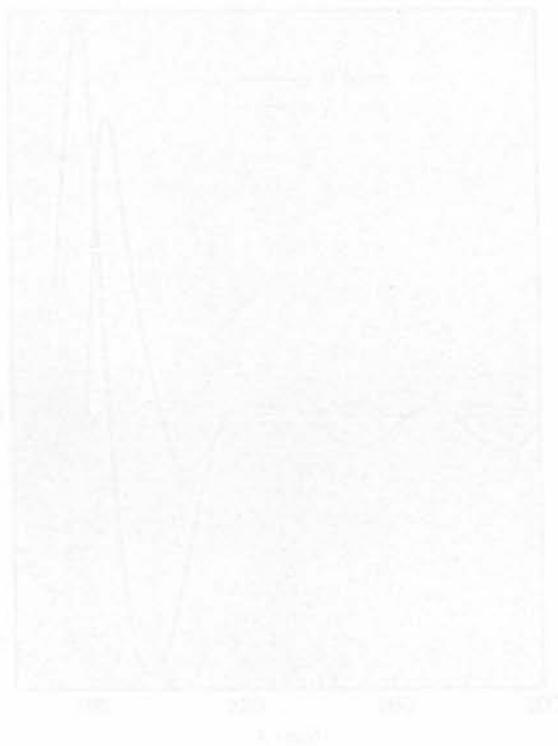


Figure 7.11: Circular Dichroism spectra (A) for protein secondary structure adapted from Greenfield and Fasman, 1969 (Greenfield and Fasman, 1969) and (B) DNA double helix adapted from Johnson, 1985 (Johnson, 1985)

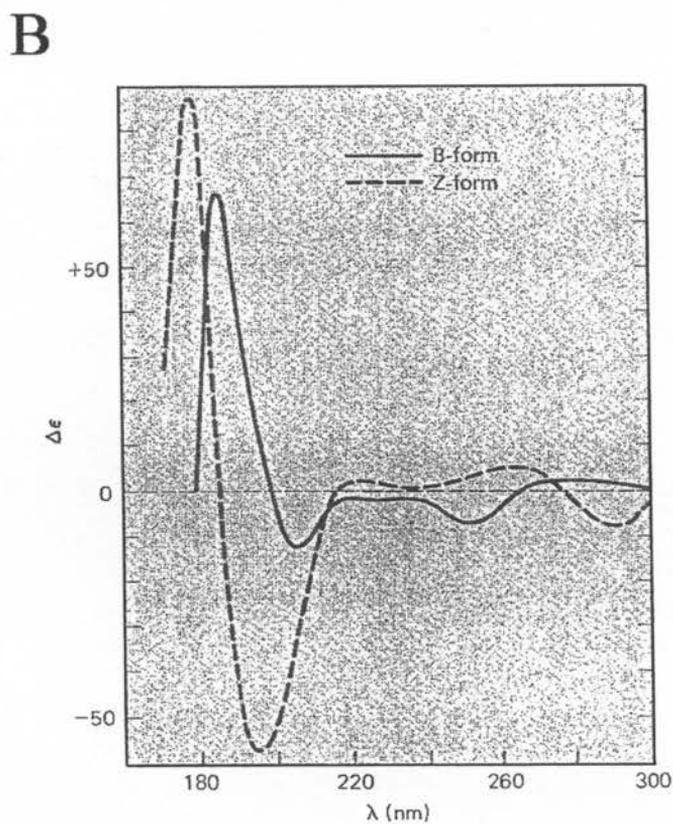
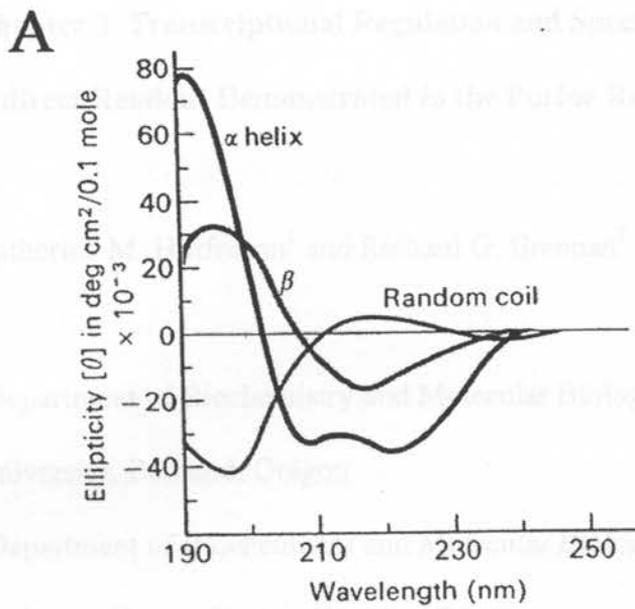


Figure 2.11: Circular Dichroism spectra A) for protein secondary structure adapted from Greenfield and Fasman, 1969 (Greenfield and Fasman 1969) and B) DNA double helices adapted from Johnson, 1985 (Johnson 1985).

## **Chapter 3. Transcriptional Regulation and Specificity for Cognate DNA Through Indirect Readout Demonstrated in the Purine Repressor (PurR) Complex**

Katherine M. Hoffmann<sup>1</sup> and Richard G. Brennan<sup>2</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, Oregon

<sup>2</sup>Department of Biochemistry and Molecular Biology, University of Texas, M.D. Anderson Cancer Center, Houston, Texas

## Abstract

The PurR-hypoxanthine-*purF* operator complex structure has been solved to 2.5 Å resolution and revealed several features key to specific DNA recognition; however, the structure does not conclusively reveal the recognition mechanism for high conservation of the Ade7 base in the operon. Equilibrium binding experiments using fluorescence anisotropy with operators containing a substitution of Ade7 by cytidine (Cyt7), thymine (Thy7) or guanine (Gua7) resulted in 12, 30 and 107 fold higher  $K_d$  values, respectively, than wild-type. The crystal structures have revealed that the mutant structures prevented proper positioning and contacts between neighboring bases and Lysine 55 in the minor groove due to changes in flexibility and packing as well as electrostatic changes in the minor groove, and in the lowest affinity operator, Gua7, and the double 2Cyt7 and 2Thy7 palendromic operators a distortion of the DNA backbone resulted of the sequence change: indirect readout. To our knowledge, this is the first systematic approach to studying the structural effects of indirect readout in a biologically relevant protein/DNA system.

## **Introduction**

Each DNA sequence has both a chemical signature characterized by the pattern of stacked base pairs exposing functional groups in the helical grooves, and a structural signature involving the local flexibility of the DNA, also dependent on the sequence of base pairs and their interaction with those immediately up and downstream. These properties allow for rapid and efficient location of a target site by a protein amidst non-specific sites and ultimately the regulation of key biological functions, such as transcription, translation, replication and recombination. (Kalodimos, Biris et al. 2004)

Both the chemical and structural signature of a given stretch of DNA is determined by the sequence of adenine, cytosine, guanine and thymine base pairs, influencing chemical repulsion and hydrogen bonding as well as structural ability to pack together, given characteristic roll, slide and propeller twisting of base pairs. Specifically, analysis of accumulated free DNA structures (el Hassan and Calladine 1996) have determined a correlation between high propeller twisting, low roll in a base pair and inflexibility of neighboring dinucleotide steps, in a broad sense determining that purine-pyrimidine steps are more flexible (a cytosine-guanine step having the highest roll and therefore being the easiest to bend) than purine-purine steps, and ultimately an adenine-adenine step, with bifurcated hydrogen bonding between neighboring bases (Nelson, Finch et al. 1987) is the least flexible dinucleotide step (el Hassan and Calladine 1996) although adenine-adenine steps, or longer adenine-tracts, may be disrupted by direct contacts with a protein (Kim, Nikolov et al. 1993; Kim, Geiger et al. 1993).

A great majority of protein DNA complex structures contain DNA that is essentially B-form, where the surface of the protein conforms to the DNA structure (Garvie and Wolberger 2001). There are, however, notable examples in which the DNA (if multiple sites, often the sequence at the distortion is conserved) is significantly deformed to accommodate protein fold,(Kim, Nikolov et al. 1993; Kim, Geiger et al. 1993; Schumacher, Choi et al. 1994; Kalodimos, Biris et al. 2004) including the LacI/GalR family of transcriptional regulators, of which the *E. coli* Purine Repressor (PurR) is a member (Rolfes and Zalkin 1988). The well characterized Lactose Repressor (LacI, (Chan, Dodgson et al. 1977; Chen, Surendran et al. 1994; Gincel, Lancelot et al. 1994; Miura-Onai, Yabuta et al. 1995; Lewis, Chang et al. 1996; Sauer 1996; Matthews and Nichols 1998; Barry and Matthews 1999)) would seem to be the ideal system for studying the mechanism by which genetic regulatory proteins discern specific target DNA sequences via structural readout: both direct, as with an electrostatic contact to a particular base, or indirect, in which the sequence of the DNA bases influences the structure of the site some distance away. The lack of low resolution structural data for LacI, however, has limited understanding of specificity of the full length protein for DNA (Lewis, Chang et al. 1996; Sauer 1996; Matthews and Nichols 1998). We believe that PurR, with it's ability to be reliably crystallized in the full length protein(Schumacher, Choi et al. 1994; Schumacher, Choi et al. 1994; Schumacher, Choi et al. 1995; Schumacher, Glasfeld et al. 1997; Arvidson, Lu et al. 1998; Huffman, Lu et al. 2002), mutants to enable otherwise prohibitively poorly bound complexes, and a functional assay that rapidly and accurately assesses protein affinity for a DNA sequence is a more appropriate system to utilize in understanding the critical role that DNA plays in

facilitating the association of DNA binding proteins and how sequence specificity and recognition is structurally accomplished on a local and global level. Certainly, it would be a welcome addition to a field that has seen few published structures pointedly demonstrating (Glasfeld, Koehler et al. 1999) indirect readout, and to our knowledge, this is the first study to demonstrate global structural implications of changing a base indirectly read by a protein (Chen, Gunasekera et al. 2001; Napoli, Lawson et al. 2006).

The *E. coli* purine repressor (PurR), a member of the LacI/GalR family, is a 76-kDa homodimer that regulates transcription of genes involved in *de novo* purine nucleotide biosynthesis and, to a lesser extent, pyrimidine nucleotide biosynthesis and salvage pathways (Rolfes and Zalkin 1988; Kilstrup, Meng et al. 1989; He, Shiau et al. 1990; Wilson and Turnbough 1990; Choi and Zalkin 1992). PurR is activated to bind cognate DNA by binding a purine corepressor, hypoxanthine or guanine, which leads to the repression of the *pur* regulon (Meng and Nygaard 1990; Schumacher, Choi et al. 1995), and regulates the transcription of at least 21 genes by binding its 16 base pair operator sites (Rolfes and Zalkin 1988). The 21 known sites comprising the *pur* operon may be described as 42 half-sites due to their pseudopalendromic nature; of these, 41 have an adenine at the seventh position, and one has a cytosine (Rolfes and Zalkin 1988; Rolfes and Zalkin 1988; He, Shiau et al. 1990; Wilson and Turnbough 1990; He, Choi et al. 1993). The structure of PurR bound to one of these operators, *purF*, as well as hypoxanthine corepressor, reveals a binding mechanism whereby Leucine 54 and its dimer-mate from the hinge helix intercalate in the minor groove of the central CpG step (by conventional numbering scheme, these are at position 9 and -9, respectively) and

bending the DNA 49-50°, and stabilizing the bend with further flexible contacts by Lysine 55 to Adenine 8. The flanking DNA is then in proper position to contact the distal HTH domains, stabilized by an interdomain contact between arginine 52 and asparagine 23, allowing for many direct electrostatic contacts between the recognition helix and the major groove, as well as water-mediated and Van Der Waals interactions. The structure did not reveal direct contacts between the adenine 7 base and the protein, although there is one set of hydrogen bonds to a water molecule that does not act as intermediary for the protein, and there are two hydrogen bonds to the phosphate oxygens of the thymine 7' base pair backbone, but no contacts to the base or sugar moiety; none of these contacts explain adenine 7's strict conservation, however (Schumacher, Choi et al. 1994).

To explore the role of the adenine at this position, and more broadly to investigate the role of well-conserved but not specifically contacted bases in DNA-binding proteins, we have conducted a systematic replacement of the adenine 7 (and its basepair, thymine 7') with cytosine, thymine and guanine and analyzed the structural as well as functional implications of these changes. Because of the pseudosymmetric nature of the *pur* operon, we have conducted these experiments on DNA mutated at single as well as double half-site positions to fully understand effect, and to compensate for statistical disorder in the crystal structures. The results reveal a variety of electrostatic changes directly affecting neighboring bases and contacts; however, we also observed significant indirect readout in each complex for which the dissociation constant was roughly 50 fold or more worse than the wild-type adenine 7 operator. These results show the striking effects of indirect readout, visually as well as empirically, in both the local and more global structure.

We present functional experiments using fluorescence anisotropy for PurR binding to *purF* operator mutated at the seventh position of one half-site and of both half-sites from adenine (wt or Ade7) to cytosine (Cyt7 or 2Cyt7), thymine (Thy7 or 2Thy7) or guanine (Gua7 or 2Gua7). We further report the crystal structures of the single half-site mutant (Cyt7, Thy7 and Gua7) and double half-site (2Cyt7 and 2Thy7) mutant operators complexed with a stabilizing mutant form of PurR and corepressor. The mutated form of PurR used in these crystallization studies is serine 53 to alanine. Serine 53 is a residue located in the hinge helix of PurR; its mutation to alanine does not disrupt the activity or fold of PurR except to favor the helical conformation associated with the bound form of the protein. This mutant form has allowed us to study PurR-DNA-corepressor complexes that might otherwise be so disfavored as to prohibit crystallization studies at low resolution.

## Materials and Methods:

*Overexpression and Purification of PurR:* Overexpressed PurR, as well as the Ser53Ala mutant, were purified as described previously (Choi and Zalkin 1992; Choi and Zalkin 1994; Schumacher, Choi et al. 1994) All protein concentrations were determined using UV absorbance at 280 nm and the extinction coefficient of PurR.

*PurR Equilibrium Binding to Operator DNA:* Fluorescence polarization experiments were done using a PanVera Beacon Fluorescence Polarization system (PanVera Corporation) as described previously (Glasfeld, Koehler et al. 1999; Huffman, Lu et al. 2002). Briefly, a 5' fluoresceinated oligonucleotide corresponding to the *E. coli purF* operator (Oligos, Etc.) (5'-F-AAAGAAAACGTTTTTCGTACCCCCTACGAAAACGTTTTCTTTT-3') in a stem-loop structure was added to 2 nM concentrations along with 250 mM potassium glutamate, 150 mM NaCl, 10 mM magnesium acetate, 1 mM ethylenediamine tetraacetic acid (EDTA), 100 mM potassium HEPES, pH 7.5, 1.0 µg/mL poly d(IC) as non-specific DNA and saturating amounts (200 µM) of hypoxanthine at 25 °C. The PurR protein was titrated into the cuvette and the sample was excited at 490 nm and emission was measured at 530 nm.

The observed millipolarization (mP) data of each binding isotherm were analyzed by least squares regression using KaleidaGraph 3.6.2 (Synergy Software, Reading, PA) according to the equation:

$$A = (A_{\text{bound}} * [P]/K_d + [P]) + A_{\text{free}} \quad (1)$$

Where  $A$  is the polarization measured at a given total concentration of PurR protein  $[P]$ ,  $A_{\text{free}}$  is the initial polarization of the free DNA, and  $A_{\text{bound}}$  is the polarization of maximally bound DNA (Lundblad, Laurance et al. 1996; Glasfeld, Koehler et al. 1999).

*Crystallization and Data Collection:* PurR protein and *purF* cognate DNA were stored, handled and crystallized as described previously (Choi and Zalkin 1992; Schumacher, Choi et al. 1994; Glasfeld, Koehler et al. 1999; Huffman, Lu et al. 2002), except that crystals of the PurR mutant-hypoxanthine-*purF* mutant operator complexes were seeded from S53A PurR-hypoxanthine-*purF* mutant or wild-type operator crystals. Briefly, a seeding crystal was harvested from its drop using a pipette in as small a volume of solution as possible, and placed in a drop of fresh protein-corepressor-operator solution pre-equilibrated and mixed 1:1 with new crystallization solution (25% PEG 4000, 0.4 M ammonium sulfate, 50 mM cobalt hexamine chloride and 0.1 M ammonium phosphate, pH 7.5). The crystal was then manually smashed and mixed by rapidly pipetting several times. 0.5  $\mu\text{L}$  of this solution of micro crystals was used to seed a new 10  $\mu\text{L}$  drop of pre-equilibrated crystallization solution, mixed, and used to serially seed the next drop. Approximately 6 serial dilutions were carried out/seed dilution, and crystals were grown using the hanging drop vapor diffusion method at 25° C. Unlike PurR complex crystals grow without seeding techniques, crystals do not initially form plates and then 3D crystals but immediately begin to grow as rhombohedron crystals, and are full size (0.8 x 0.5 x 0.4) in a matter of weeks. The resulting crystals are isomorphous to wild-type PurR-hypoxanthine-*purF* operator crystals, and diffract comparably. X-ray intensity data for all crystals were collected at 22° C on an ADSC area detector using a Rigaku RU-200

X-ray generator at 40 kV, 150 mA. The data were processed using the D\*Trek software (Pflugrath 1999), except G7, which was processed with BIOTEX (Molecular Structure Corporation, Inc. Woodlands, TX).

*Structure Determination and Refinement:* Structures were solved using difference Fourier techniques utilizing the wild-type PurR-hypoxanthine-purF operator structure (1QPZ) as the initial phases and model for each of the holo complexes, after removing all solvent. The asymmetric unit contains one PurR monomer, one purine base and a *purF* operator half-site.

Refinement of each complex was initiated with rigid body refinement, followed by XYZ and B (thermal) refinement, as implemented in TNT ((Tronrud 1997), used initially for the single half-site operator mutants) or CNS ((Brunger, Adams et al. 1998), for the double half-site mutants and for the single half-site final refinement as well, for consistency). Refinement was monitored, and the structure manually adjusted using the program XtalView (McRee 1999). Fo-Fc omit maps were calculated for each structure to ensure the unbiased placement of side chains, operator bases and solvent molecules. The stereochemical quality of each structure was ascertained using PROCHECK (Laskowski, Rullmann et al. 1996), RMSDs calculated using LSQKAB program (Kabsch 1976) for alpha carbons from residue 2 to 340 in the A chain, distances less than 5 Å between molecules calculated using the program CONTACT, all in the CCP4 suite of programs (1994). Analysis of DNA structure and bend was done with the program CURVES 5.2 (Sklenar, Etchebest et al. 1989).

## Results:

Crystals were grown of S53A PurR bound to *purF* operator substituted at the adenine 7 position with cytosine, thymine and guanine. Functional assays using fluorescence anisotropy with the same protein and substitutions in the *purF* sequence indicated dissociation constants 12, 30 and 107 times higher than to the wild-type operator. (Table 3.02) The structures of the single half site mutants offered some insight into the structural preference for adenine at position 7; however, there was significant statistical disorder around position 7, due to the pseudopalindromic nature of the binding site, and studies were also conducted on double half site mutants of a truly palindromic operator to resolve this ambiguity (Table 3.01). Functional studies on the double half site substituted cytosine, thymine and guanine at the 7 position in a palindromic operator indicated dissociation constants 115 and 322 fold greater than wild type for cytosine (2Cyt7) and thymine (2Thy7), and no binding was detected to the double substituted guanine operator (Table 3.02), nor were crystals of this complex ever observed.

Overall quaternary, tertiary and secondary structure is maintained in all structures, RMSDs uniformly between .214 and .304 Å with maximum variations at discrete locations. Protein contacts to the phosphate backbone at the 7' base are maintained in all structures except 2Thy7, where each hydrogen bond is weakened past 3.4 Å, and additionally in both Cyt7 and 2Cyt7 between N23 and the O1P of the 7' base. The wild-type contacts between the flexible K55 and positions 9, 8 and 7, however, vary in each of the structures: Cyt7 loses the contact to Cyt9, Thy7 loses that and the contact to Ade7' as well, Gua7 maintains only the contacts to the 7' base, and gains two hydrogen bonds to

the Guanine7 base as well, the double mutant structures lose the contacts to the 7' base, but maintain the contacts to Cyt9/Gua9', in contrast to the single mutants. (Table 3.03)

Global DNA structural comparison reveals that Cyt7 and Thy7 maintain the wild-type 49° kink at the central CpG step throughout the binding site. Gua7, 2Cyt7 and 2Thy7, however, do not maintain the wild-type bend, instead displaying global bends of 42, 45 and 40° in correlation with dissociation constants more than 100 fold worse than wild type (table 3.02, figure 3.02). More localized analysis of DNA structure reveals increasingly relaxed propellor twist at the 7-7' base with increasing  $K_d$ . Local inter-base pair roll is decreased by more than 27% in the double mutant structures, particularly 2Cyt7, where the decrease is 63% between position 7 and 8, and 81% between positions 6 and 7; 2Thy7, in contrast, has a 27% decrease in roll between 7 and 8, but makes up for it with a 59% increase between 6 and 7. Inter-base pair twist between positions is consistent with wild-type in both Cyt substitutions, but Thy7 shows a decrease (12%) between positions 6 and 7, 2Thy7 showing a marked decrease (39%) and additional marked increase (32%) between positions 5 and 6. (Table 3.02, figure 3.03)

### **Cyt7**

The single half site cytosine substitution reflects the only known natural substitution for adenine at this position in the PurR operon, and also had the binding activity most similar to the wild type adenine 7 DNA (table 3.02, figure 3.01b). The structure of the single cytosine 7 substituted DNA/PurR complex was similarly consistent with the wild-type structure: there were no significant global structural differences between the two protein structures, and an overlay of the oligonucleotide structures at the phosphate backbone

revealed an RMSD of 1.4 Å (table 3.03, figure 3.01b). Closer examination of the 7 position, as well as its nearest neighbors (adenines 6 and 8) reveals several small changes relative to the wild type structure (1qpz in the protein data bank, (Schumacher, Choi et al. 1994; Glasfeld, Koehler et al. 1999)). Most prominently, the protein residue lysine 55, which in the wild type structure makes hydrogen bonds to the adenine 8 in the minor groove, is in the cytosine 7 structure making weak hydrogen bond contacts to the N3 and C2 amino group of the guanine 7' base pair (table 3.03). Previous structures focusing on K55 position (Glasfeld, Koehler et al. 1999) have indicated a fair amount of flexibility in the position of this residue, allowing it to contact adenine 8 preferentially, but also to swing out of position and contact the 7' base pair (thymine in the wild type structure). Nevertheless, the electrostatic differences in the DNA minor groove between an AT and a CG base pair (a significant increase in positive character in the prime base as the oxygen functional group at C2 on thymine is replaced by an amino C2 group of guanine), as well as the size difference between thymine and guanine may alter the environment of the minor groove such that K55s preferential position is now nearer the 7 position base pair (figure 3.03b). This structural difference could account for the difference in binding to a site containing a cytosine at position 7, as the behavior of K55 has influenced binding constants in other PurR studies (Glasfeld, Koehler et al. 1999). There is additionally a structurally looser stacking of cytosine 7 between adenines 8 and 6 than in the string of adenines found in the consensus sequence. Purine strings have been shown to form more rigid and tightly packed DNA structures than purine-pyrimidine steps, due to the ability to pack closer together. Adenines more specifically form particularly rigid rod structures when found adjacent to each other: this seems to be a

result of the AT basepairs high propellor twist, and the ability of an adjacent pair to twist similarly and pack very tightly and resist sliding (el Hassan and Calladine 1996; Dickerson 1998). The introduction of a pyrimidine (cytosine) in the center of the wild type adenine string interferes with the tight packing of the bases; however, the DNA seems to compensate for this by adjusting the position of the adenine at position 8, which is rotated away from the minor groove and may further influence K55 to preferentially contact the guanine in the 7' position. The positive electrostatics and larger bulk of the guanine 7' may also motivate the shift in adenine 8, though the relatively small difference in binding constant and the presence of the cytosine substitution in the PurR operon indicates that the adjustment of adenine 8 and K55 positions are relatively mild obstacles to binding.

### **Thy7**

The thymine half site mutant shows structural changes similar to that of the cytosine structure, but slightly more exaggerated: K55 is still in a position to contact the 7' position rather than the adenine 8, however, in the thymine structure, it is further away from the adenine 8 in order to tightly hydrogen bond with the N3 of adenine 7' (base pair to the thymine substitution, shown in figure 3.03c, reported in table 3.03). The strong propellor twist of the AT base pair is twisted in the opposite direction in the thymine 7 structure compared to the wild type, and this twist seems to require more space between neighbors, as adenine 8 is not only rotated to compensate, but is also shifted away from thymine 7. This increase in space between bases is compounded by the removal of bulky amino groups present in the guanine 7' base in the cytosine structure compared to the thymine 7-adenine 7' basepair, allowing for even more flexibility in K55. All of this is

consistent with the anisotropy experiments showing a decrease in affinity for Thy7 beyond that of Cyt7, as well as the lack of thymine substitutions in the *pur* operon (table 3.02, figure 3.01b). The cytosine versus thymine structures, while similar in nature, represent the difference between a sequence the protein is able to bind, and a sequence character different enough that PurR cannot compensate *in vivo* (where time and concentration will be limiting,) however, we felt that greater resolution, particularly in the area around position 7, where there was some statistical disorder, was needed to confirm these observations for both substitutions.

### **2Cyt7 and 2Thy7**

The double half site mutant DNAs for cytosine and thymine have PurR dissociation constants 115 and 322 fold worse than the wild type operator; this roughly 10 times worse than the single half site mutants, and significantly higher than the minimum expected effect of 2 times the single half site constant (table 3.02, figure 3.01c). The double cytosine mutant, interestingly, showed K55 in nearly the same position as the half site model (figure 3.03b). The double thymine model, however, placed K55 in a nearly wild type position contacting adenine 8 (figure 3.03c). The near wild type K55 behavior in the 2Thy7 structure is particularly notable since the backbone of the DNA for both double half site mutants is significantly distorted, compared to the wild type and single DNA mutants, indicating that indirect readout effects of the double mutation may prohibit efficient binding by PurR (figure 3.01, b and c). This significant structural difference, along with the poor binding constant, implies that the K55 flexibility indicated in the single half site model may have allowed the protein to make the wild type like contact as a requisite for binding the double half site mutants under *in vitro* conditions.

The additional contact may have been what allowed the protein to form the complex with the substituted DNA given that it was not necessary in the single half site thymine mutant, nor either of the cytosine structures. Both bases flanking the substituted 7 positions in the double site mutant structures are also significantly altered (figure 3.03b and c), while in the single site mutants the 6 position seemed to be relatively stable. This is concordant with the significant backbone distortion about the 7 position, probably due to stacking differences between purine-pyrimidine steps and purine-purine steps. This indirect readout effect was not observed in the single half site mutant structures, possibly due to the proteins ability to force them into favorable conformation when more tightly bound, or possibly due to statistical disorder; however, the indirect readout due to flexibility does explain the significant functional differences in PurR's preference for a purine string rather than a purine-pyrimidine sequence (el Hassan and Calladine 1996; Dickerson and Chiu 1997; Dickerson 1998), implied in the dramatically different binding affinity.

### **Gua7**

In contrast to the mild structural and functional differences seen in the cytosine and thymine single half site DNA mutants, the guanine substitution at the 7 position introduced a significant impediment to binding, as evidenced by >100 fold increase in dissociation constant of binding, on a level with the double mutant structures. The tertiary complex structure with Gua7 DNA reveals a similarly larger structural difference, 42° rather than a 49° bend in the bound state reflecting indirect readout. On closer observation, the major reason for this seems to be a relief of the bend at the substituted position. Where the pyrimidine substitutions (cytosine and thymine) seemed to introduce

flexibility in the DNA that PurR was able to compensate for (arguably in the double mutants, but adequately in the single site mutants,) guanine substituted for adenine at the 7 position seems to force the DNA into an incompatible structure that is nevertheless relatively rigid. While adenine strings are particularly known for rigidity, they are also characterized by a high degree of propeller twist between base pairs and tight packing between neighboring bases (el Hassan and Calladine 1996). A guanine substitution, on the other hand, is unable to pack sufficiently tightly, establishing a rigid and incompatibly smaller propeller twist with its cytosine base pair; introducing significant difficulty for the protein attempting to bend it into place and make critical flanking contacts.

The nature of the guanine-cytosine base pair also introduces a mild chemical difference to the minor groove which influences the behavior of K55 and seems to attract it into a closer hydrogen bond with the 7' base pair, and for the first time to a 2.67Å hydrogen bond with the seventh position. This is in contrast to the cytosine and thymine substitutions, which likely due to electrostatics as well as the size change of purine for pyrimidine, repel K55 away from contacts with the 7' base and into weaker contacts with position 8 base pair (table 3.03).

## Discussion

### Geometry vs. Energetics

Our five DNA substituted structures indicate disparate causes for decreased affinity exemplifying both direct and indirect readout. The DNA containing substituted bases least disruptive to protein binding (Cyt7 and Thy7) caused relatively mild disruption of protein-DNA contacts in the hinge helix, and little distal or global disruption at all. This could be due either to the proteins ability to compensate for the change (likely in the case of the single Cyt7 mutant since this substitution is the only one found in the known operon) or the crystallographic statistical disorder. The doubling of the Cyt7 and Thy7 to double half-site substitutions (2Cyt7, 2Thy7) significantly deteriorated the ability of the protein to successfully create a stable protein-DNA complex, manifested in decreased protein-DNA contacts at the hinge helix (K55) combined with structural difficulty achieving the full bend in the DNA. This combined effect had the most profound effect on affinity of the five structures solved. The Gua7 structure is differently disruptive, as the guanine base forces a DNA half-site structure that is not only less preferred, but is incompatible with the formation of a stable complex where the other substitutions introduced structural flexibility that the protein could compensate for in limited amounts.

In contrast to the increase in conformational mobility seen in the pyrimidine substitutions, the guanine for adenine substitution leads the protein to adopt shorter hydrogen bonds between K55 and the DNA, but also to different bases, possibly supporting the relief in global bend; a less propeller twisted conformation combined with additional functional groups in the minor groove mean the Gua7 base structurally prohibits the straight, tightly packed bound structure and further doesn't allow for

recovery the way the pyrimidine substitutions do, as a result, the structural relief of bend is translated down the binding site leading to a global relief of DNA bend. There may still be statistical disorder blurring the details of the single guanine substituted mutant structure, particularly as regards K55 contacts, and a double half-site mutation to guanine might have clarified the details of the decreased affinity for this substitution; however, the double half-site mutation to guanine has no measurable binding and will not form crystals. Gua7 showed dissociation constant and DNA bend on a similar scale as the double 2Cyt7 and 2Thy7 substitutions; we suspect that the reason no binding of the double 2Gua7 substituted DNA by the protein was observed, and no crystals could be obtained, is due to an amplification of the indirect readout, and possibly a larger structural change. We conclude that the guanine base introduces a different DNA geometry where cytosine and thymine introduce greater flexibility that the protein was able to compensate for in the single substitutions. In the biologically relevant single cytosine substitution, the protein was able to completely compensate and attain wild type contacts and structure, though the binding constant indicates a slightly greater energetic requirement for making the compensation. The difference in binding observed with the thymine and cytosine structures therefore involved a greater energetic requirement for binding, where the guanine structure demonstrates a geometric barrier to binding.

### **Comparison to dissociation models for LacI**

Our inability to measure a dissociation constant for the 2Gua7 DNA or attain diffraction quality crystals led us to wonder if PurR is unable to make the coil to helix transition in the hinge helix or possibly if disruption of proper HTH contacts was preventing

crystallization of the tertiary complex (the coil to helix transition in PurR triggered by DNA binding presents a structural change required for crystallization of the full length protein (Schumacher, Choi et al. 1994; Schumacher, Choi et al. 1995)). Exploration of the literature led us to consider experiments studying the dissociation of LacI from its promotor. LacI and PurR have 35% sequence identity in their monomers and highly homologous secondary and tertiary structures (Rolfes and Zalkin 1988; Weickert and Adhya 1992; Chuprina, Rullmann et al. 1993; Schumacher, Choi et al. 1994; Friedman, Fischmann et al. 1995; Nagadoi, Morikawa et al. 1995; Nagadoi, Nakazawa et al. 1995; Lewis, Chang et al. 1996; Spronk, Bonvin et al. 1999; Spronk, Folkers et al. 1999; Bell and Lewis 2000). While there are some striking structural and functional differences between the two proteins in the core domain, notably regarding effector binding, there are nevertheless very similar allosteric conformational changes in the DNA binding domains involving the coil to helix transition in the 10 amino acid hinge region linking the HLH DNA binding motif to the rest of the protein (Schumacher, Choi et al. 1995; Lewis, Chang et al. 1996; Matthews and Nichols 1998; Mowbray and Bjorkman 1999; Bell and Lewis 2000). Since many of the structures regarding the dissociation of LacI from DNA were done using a truncated version of the protein encompassing the DNA binding domain and linked hinge, comparisons to PurR are highly relevant (Swint-Kruse, Larson et al. 2002).

Deuterium exchange experiments utilizing a disulfide bond linked DBD of LacI

(Kalodimos, Biris et al. 2004) indicated a mode of dissociation involving first a helix to coil transition in the hinge helix of LacI, followed by more universal dissociation of the

HLHs from DNA; our indirect readout structures might also evidence a destabilization of the hinge helix. We found, however, that all contacts between the DBD and DNA are consistent with the wild type structure, although the 2Thy7 structure evidences strained bonds; in the R52-7'O1P connection, for example, an approximately 3.0Å hydrogen bond is maintained except for the 2Thy7 structure, which has a 3.4Å bond. The K55 hinge helix residue seems to have base specific influences as discussed in the text, and the key L54 residue intercalating between the central base pairs makes connections consistent with the wild type structure.

Further examination of hydrogen bonding contacts between the protein and the binding site reveal, however, a number of contacts between the HTH residues and the DNA that are disrupted in the 2Thy7 structure (S19-7'O2P at 3.63Å, and N23-7'O1P at 4.39Å are reported in table 3.03), and one that is also weak in 2Cyt7 (N23-7'O1P at 3.56Å); notably, the bidentate contact between R26 and Gua2 that is 2.7 and 3.07Å in the wild type structure, is 3.05 and 3.88Å in the 2Thy7 structure. Our data indicates that while the helix to coil transition may be an early structural evidence of dissociation from the DNA in truncated domain experiments, we see relief of DNA bend, followed by disruption of the stabilizing contacts to the HTH that may precede the unwinding, all of which seems to precede disruption of electrostatic contacts to the hinge helix in the full length protein complex.

## **Conclusion**

We have conducted structure-function studies on the DNA binding protein PurR bound to cognate DNA substituted at a well conserved, but non-specifically contacted base

(Adenine 7) in the sequence. The structures revealed that the electrostatic differences inherent in base substitutions influence neighboring base-protein contacts, that base stacking abilities of different base pairs could affect local DNA structure, and that these disturbances, if amplified, may result in global changes to the DNA structure (indirect readout) sometimes resulting in disruption of protein-DNA contacts distal to the substitution. We suggest these crystallographic structures represent snapshots, not only of indirect readout and the effects of base pair changes in the binding site, but may also represent the protein in various stages of association with DNA, allowing us to hypothesize a mechanism of dissociation.

Figure and Table Legends:

Table 3.01: Selected Crystallographic Data and Statistics.

Table 3.02: DNA Structural Analysis and Selected Structural and Functional Statistics.

Table 3.03: Analysis of Protein Contacts (Hydrogen Bond Distances).

Figure 3.01: Comparative Binding Isotherms for PurR binding to the single and double half-site variants of *purF*. The data are an average of three experiments with the error bars indicating one standard deviation from the average; average, standard deviations and dissociation constants calculated with the program KaleidaGraph 3.6.2. A) wt PurR binding to wt *purF* (purple) and S53A PurR (black), the stable mutant used in the remainder of these binding experiments and in the structures, B) S53A PurR binding to single half-site mutants Cyt7 (blue), Thy7 (green) and Gua7 (red), C) S53A PurR binding to double half site mutants 2Cyt7 (blue) and 2Thy7 (green), no results were obtained for 2Gua7 binding. Dissociation constants are reported in Table 2.

Figure 3.02: A) Overlay of full-length PurR/hypoxanthine/*purF* structure (color coded DNA) on the S53APurR/hypoxanthine/Gua7 structure (red DNA) B) overlay of the DNA binding domains and binding sites for the S53APurR/hypoxanthine/Cyt7 and S53APurR/hypoxanthine/2Cyt7 structures, light and dark blue DNA, respectively, C) overlay of the DNA binding domains and binding sites for the

S53APurR/hypoxanthine/Thy7 and S53APurR/hypoxanthine/2Thy7 structures, in light and dark green DNA, respectively.

Figure 3.03: Closeup on the substituted DNA bases overlaid on the wild-type Adenine. Structural variations from the wild-type are indicated, and positions of protein residues from both the single and double site mutants, if applicable, are shown as sticks. A) Gua7 in pink atom specific colors, with highlighted relief of propellor twist, and increase in interbase twist compared to wild-type. B) 2Cyt7 in blue atom specific coloring with highlighted relief of propellor twist and increased interbase roll. C) 2Thy7 in green atom specific coloring, also with highlighted relief of propellor twist and increased interbase roll.

Table 3.01. Selected Crystallographic Data and Statistics.

A. Data Collection	Cyt7	Thy7	Gua7	2Cyt7	2Thy7
Space Group	C222 <sub>1</sub>				
Cell constants: a=	175.9	175.65	174.49	175.90	175.46
b=	95.13	95.40	95.25	95.13	95.10
c=	81.54	81.52	82.02	81.56	81.98
Unique reflections	19,402	24,122	22,954	21,434	20,000
Total reflections	69,097	87,350	61,520	77,091	77,576
Completeness (%)	100.0	100.0	85.4	94.7	97.0
Overall R <sub>sym</sub> <sup>a</sup>	14.1	7.0	9.1	5.1	7.0
I/σ(I) - all data	3.5	6.3	6.0	3.2	2.2

## B. Refinement

Resolution limits (Å)	30.0 – 2.69	30.0 - 2.50	30.0 - 2.54	30.0 - 2.65	30.0 – 2.60
R <sub>work</sub> (%) <sup>b</sup>	22.9	21.4	20.7	20.0	20.7
R <sub>free</sub> (%) <sup>b</sup>	27.4	24.7	25.9	23.2	23.1
Number of atoms	3055	3068	3085	3118	3092
Solvent molecules	42	56	83	103	77

## C. Stereochemistries

rms bond dist. (Å)	.008	.008	.014	.007	.007
rms bond angles (°)	1.14	1.13	1.38	1.05	1.09

<sup>a</sup>  $R_{sym} = \frac{\sum \sum |I_{hkl} - I_{hkl(j)}|}{\sum I_{hkl}}$ , where  $I_{hkl(j)}$  is the observed intensity and  $I_{hkl}$  is the final average intensity value.

<sup>b</sup>  $R_{work} = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$  and  $R_{free} = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$ , where all reflections belong to a test set of 10% randomly selected data.

Table 3.02: DNA Structural Analysis

	Ade7	Cyt7	Thy7	Gua7	2Cyt7	2Thy7
Dissociation Constant (Kd, nM, error in parentheses)	2.5(.7)	15(2)	100(16)	270(30)	290(40)	800(80)
Global Bend (P-P, degrees)	49	50	48	42	45	40
Propellor twist at 7-7'	-11.8	-8.0	-8.7	-5.3	1.9	2.6
Local interbase-pair roll ( $\rho$ ) Between 7 position and A8	5.66	5.72	6.04	6.07	2.09	4.16
Local interbase-pair roll ( $\rho$ ) Between A6 and 7 position	2.04	2.16	1.73	2.51	-0.64	3.25
Local interbase-pair twist ( $\omega$ ) Between 7 position and A8	37.70	34.91	33.49	29.79	39.94	37.74
Local interbase-pair twist ( $\omega$ ) Between A6 and 7 position	35.81	34.51	31.21	40.84	38.95	21.83
Local interbase-pair twist ( $\omega$ ) Between C5 and A6	36.61	33.27	35.73	36.57	34.73	48.45

Table 3.03: Analysis of Protein Contacts and Statistics.

	Ade7	Cyt7	Thy7	Gua7	2Cyt7	2Thy7
RMSD <sup>a</sup> ave /maximum	0/0	.299/1.13	.232/.945	.304/.825	.214/.670	.238/.956
R26-Gua2	2.7/3.07	2.96/3.30	3.00/3.11	2.76/3.25	2.98/2.78	3.05/3.88
N23-7'O1P	3.20	3.46	3.06	3.17	3.56	4.39
R52-7'O1P	2.86	3.04	3.14	2.72	2.98	3.44
S19-7'O2P	2.96	2.43	2.69	2.88	3.17	3.63
K55-7base	-	-	-	2.67/3.31	-	-
K55-7'base	3.42/3.27	3.29/3.30	-	3.29	-	-
K55-Ade8/Thy8'	2.81/3.31	2.90/3.43	3.16/3.43	-	3.17/3.48	3.23/3.38
K55-Cyt9/Gua9'	3.31	-	-	-	3.33	3.30

<sup>a</sup>RMSD is root-mean-square deviation calculated for C $\alpha$  atoms.

Figure 3.01:

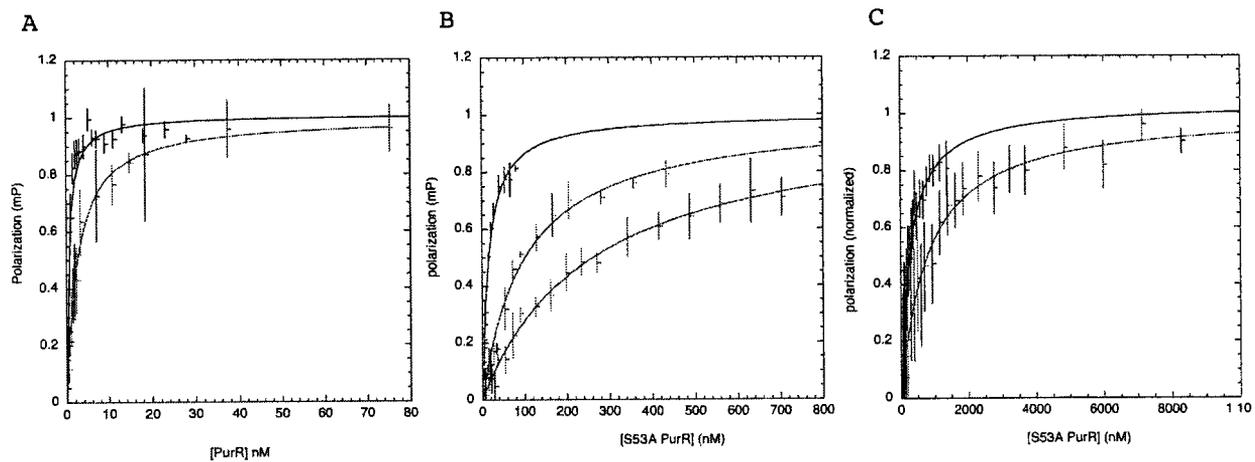


Figure 3.02:

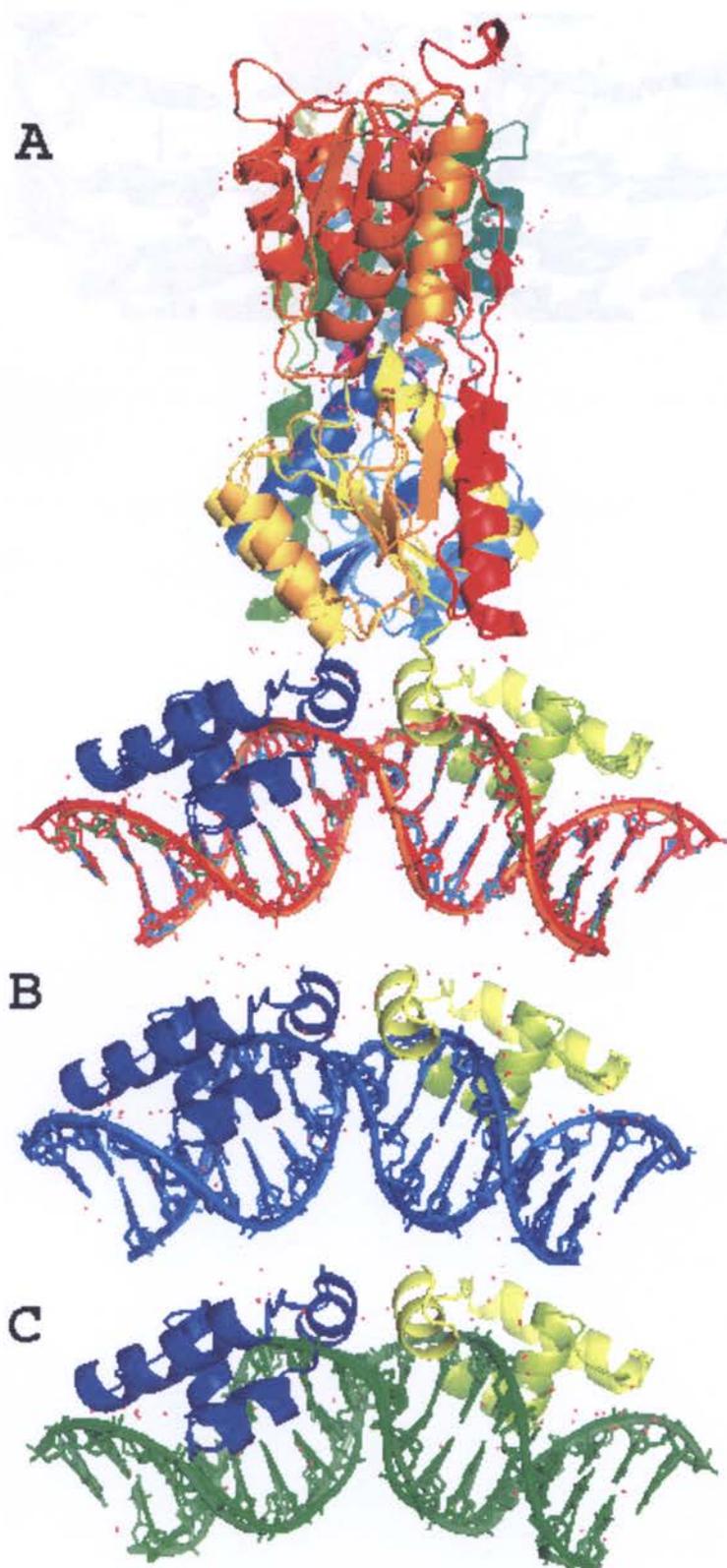
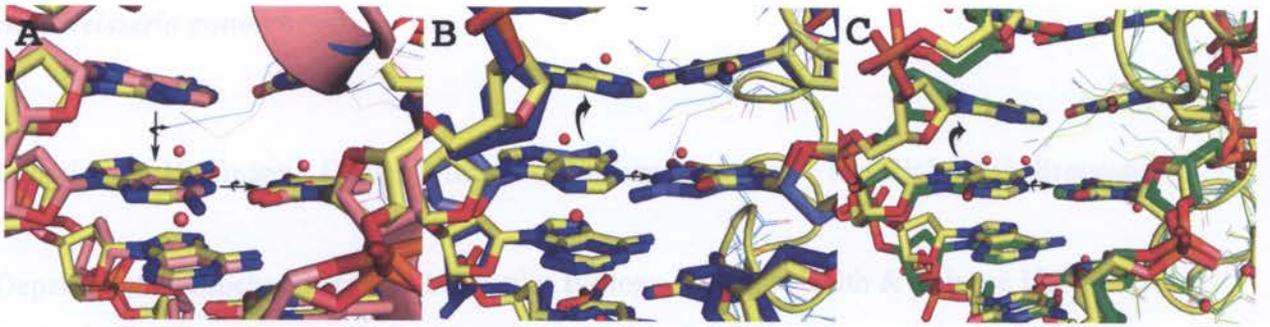


Figure 3.03:



Department of Microbiology and Immunology, Oregon Health Sciences University, 3181 S.W. Sam Jackson Park Road, Portland, Oregon 97239-3098  
Laboratory of Molecular Pathogenesis, VA Medical Research Service, Veterans Affairs Medical Center, Durham, North Carolina

**Chapter 4: Characterization of the Multiple Transferable Resistance Repressor, MtrR,  
from *Neisseria gonorrhoeae***

Katherine M. Hoffmann<sup>1</sup>, Daniel Williams<sup>2</sup>, William M. Shafer<sup>2,3</sup> and Richard G. Brennan<sup>1\*</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Oregon Health & Science University,  
Portland, Oregon

<sup>2</sup>Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta,  
Georgia

<sup>3</sup>Laboratories of Microbial Pathogenesis, VA Medical Research Service, Veterans Affairs  
Medical Center, Decatur, Georgia

## **Abstract**

MtrR represses expression of the *Neisseria gonorrhoeae* *mtrCDE* multidrug efflux transporter genes. MtrR displays salt dependent DNA binding, a stoichiometry of two dimers per DNA site and, for a protein that was expected to be essentially all helical, a high percentage of random coil and possibly  $\beta$  sheet structure.

## Introduction

In order to colonize human mucosal membranes, *Neisseria gonorrhoeae* (GC) must overcome host defense mechanisms that include exposure to potentially lethal levels of antimicrobial hydrophobic agents. Early studies implicated the multiple transferable resistance (*mtr*) locus as a key determinant in resistance that initially was thought to play a role in the modification of the gonococcal cell envelope during colonization (Guymon, Walstad et al. 1978). Later studies demonstrated that this locus encoded a three gene operon, designated *mtrCDE*, which forms an energy dependent efflux system that expels multiple hydrophobic agents (Hagman, Pan et al. 1995; Delahay, Robertson et al. 1997). The MtrD protein is a multidrug efflux transporter that belongs to the resistance/nodulation/division transporter family (Hagman, Lucas et al. 1997). The MtrC protein belongs to the membrane fusion protein family that links MtrD with MtrE, an outer membrane protein, which serves as a channel for export of antimicrobials to the extracellular environment (Delahay, Robertson et al. 1997).

Due to the broad substrate specificity of multidrug efflux transporters, which could result in the accidental efflux of needed metabolic intermediates, the expression of their genes is regulated tightly (Grkovic, Brown et al. 2002). Transcription of the *mtrCDE* operon is controlled by both *cis* and *trans* acting factors under the influence of the *mtrR* gene (Hagman and Shafer 1995). Missense or deletion mutations of the *mtrR* gene in clinical isolates leads to increased transcription of *mtrCDE* and a consequential increase in antimicrobial resistance, thus confirming the repression of MtrCDE efflux pump transcription by MtrR (Veal, Yellen et al. 1998). The *mtrR* gene is located ~250 bp upstream of the *mtrCDE* genes and is transcribed divergently from that operon. The *mtrR* gene encodes a 210 amino acid residue, ~23 kDa protein

(MtrR). MtrR contains a putative N-terminal helix-turn-helix (HTH) motif and amino acid sequence similarity to several members of the TetR/CamR family, e.g., 53% identity, 78% homology to AcrR (Pan and Spratt 1994; Aramaki, Yagi et al. 1995; Lucas 1997; Stapleton, Adams et al. 2004). Footprinting and DNaseI protection experiments established that MtrR protects a 22 to 27 base pair region upstream of *mtrC* (Lucas 1997). Although a DNA binding site was identified and the DNA binding domain bears strong homology to those of other TetR family members, the stoichiometry of MtrR binding to DNA is unknown. TetR family members have shown variability in their DNA binding oligomerization states (Hillen and Berens 1994; Grkovic, Brown et al. 2001; Schumacher, Miller et al. 2002; Engohang-Ndong, Baillat et al. 2004).

In order to understand better the DNA binding properties of MtrR and its role in GC resistance against hydrophobic agents and other antibiotics, we carried out a biophysical and biochemical characterization of this multidrug efflux pump gene repressor. These studies included the determination of the length of cognate DNA required for optimal MtrR binding, the effect of NaCl concentration on DNA binding affinity, the stoichiometry of binding and its secondary structure in the presence or absence of cognate DNA. Unanticipated differences were observed between the MtrR and TetR family member, QacR, the *Staphylococcus aureus* multidrug binding transcription repressor (Grkovic, Brown et al. 1998).

## Methods and Results

**Cloning, expression and purification of MtrR.** The 633 base pair *mtrR* gene from *N. gonorrhoeae* strain FA19 was PCR amplified from chromosomal DNA using primers that contained engineered restriction sites *NdeI* and *BamHI*. After digestion with *NdeI* and *BamHI*, the fragment was cloned into a pET-15b ampicillin-resistant vector containing an N-terminal hexahistidine affinity tag followed by a thrombin cleavage site. The vector was sequenced to ensure fidelity and transformed into Rosetta-gami B(DE3)pLysS cells containing resistance to chloramphenicol. One litre cultures were grown in Luria-Bertani broth containing 100 µg/mL ampicillin and 50 µg/mL chloramphenicol at 37 °C to an OD<sub>600</sub> of 0.6 AU, at which time cells were induced with 1 mM IPTG for 3 hours. Cells were then centrifuged and resuspended in 20 mM Tris pH 7.6, 500 mM NaCl, 10% glycerol and 1mM Tris(2-Carboxyethyl)Phosphine Hydrochloride (TCEP), as a reducing agent. The cells were lysed by French Press after which the lysate was centrifuged and the supernatant loaded onto a Ni<sup>2+</sup>-NTA column. Pure hexahistidine-tagged MtrR was eluted with Buffer A (100 mM Na<sup>+</sup>/K<sup>+</sup> phosphate buffer, pH 8.5, 300 mM NaCl, 5% glycerol and 1mM TCEP) containing 500 mM imidazole (data not shown). Fractions were analyzed by Q-TOF mass spectrometry and SDS-PAGE before dialyzing overnight into 200 mM Na<sup>+</sup>/K<sup>+</sup> phosphate pH 7.5, containing 10% glycerol and 1mM TCEP (Phosphate Storage Buffer, PSB). Specific and complete cleavage of the hexahistidine tag was attained. However, the stability of MtrR was compromised and therefore, only purified His-tagged MtrR was used in our studies.

**DNA Binding Affinity and Binding Stoichiometry.** A fluorescence polarization based assay was used to determine the DNA binding affinity of MtrR for a pair of oligodeoxynucleotides

from the *mtrCDE* promoter. These oligodeoxynucleotides were purchased from Oligos Etc. (Wilsonville, OR) and were 27 and 22 base pairs with fluorescein covalently attached to their 5' end by a hexamethylene linker. Each oligodeoxynucleotide encompassed the direct repeat that in previous footprinting studies were protected to different extents by MtrR. The respective sequences of one strand of the 27mer and 22mer are

5'-TTTTTATCCGTGCAATCGTGTATGTAT and 5'-ATCCGTGCAATCGTGTATGTAT with the pseudo direct repeats in bold (Lucas 1997). The standard DNA binding solution used in these studies was 20 mM Tris-HCl pH 7.4, 100 mM NaCl, 0.1 nM fluoresceinated DNA (or higher, provided  $K_d \geq 10 \times$  the DNA concentration), and 1  $\mu$ g of poly d(IC), as nonspecific DNA. MtrR in PSB was titrated into the binding mixture until the millipolarization no longer rose. All experiments were carried out at 25 °C. The excitation wavelength was 490 nm and units of fluorescence polarization (millipolarization) were read at 530 nm. The data were plotted using the equation,  $P = \frac{(P_{\text{bound}} - P_{\text{free}})[\text{Protein}]}{K_d + [\text{protein}]} + P_{\text{free}}$ , where P is the polarization measured at a given total protein concentration,  $P_{\text{free}}$  is the initial polarization of free fluorescein-labelled DNA, and  $P_{\text{bound}}$  is the maximum polarization of specifically bound DNA.  $[\text{Protein}]_{\text{free}} = [\text{Protein}]_{\text{total}}$  is assumed because the concentration of fluorescein-labelled DNA is 10 fold below the  $K_d$ . The generated hyperbolic curves are fit by nonlinear least squares regression analysis, assuming a bimolecular model such that the  $K_d$  values represent the protein concentration at half-maximal ligand binding and plotted using the graphing program Kaleidograph (Lundblad, Laurance et al. 1996). The longer oligodeoxynucleotide ( $K_d = 0.9$  nM) bound ~9 fold better than the shorter oligodeoxyoligonucleotide ( $K_d = 7.8$  nM) (Figure 1A). Increasing the oligodeoxynucleotide length to 31 base pairs did not result in higher affinity (data not shown). Consequently, the remaining DNA binding experiments used the 27mer.

The length of the higher affinity DNA binding site of MtrR is nearly identical to the high affinity DNA binding site (IR1) of QacR. Two dimers of QacR bind the 28 base pair IR1, which is located in the promoter region of the *qacA* multidrug efflux pump gene, and although pseudopalindromic, IR1 contains four pseudo direct repeats that interact with QacR (Grkovic, Brown et al. 2001; Schumacher, Miller et al. 2002). To determine whether MtrR employs the same or a different stoichiometry of binding to the *mtrCDE* promoter, a fluorescence polarization assay was utilized. The binding buffer and conditions were the identical to those used in the binding affinity determination experiments except that the concentration of the 27mer was increased to 20 nM, i.e., > 20 fold above the  $K_d$ , thereby ensuring stoichiometric binding. MtrR was titrated into the binding solution in until the total protein concentration (in monomers) reached 200 nM. The graph of the resulting data shows a linear increase in the observed millipolarization until saturation of the high affinity DNA sites, after which low affinity DNA binding takes place (Figure 2). The inflection point occurs at an MtrR monomer concentration of 80 nM, which when divided by the concentration of cognate DNA (20 nM), indicates a stoichiometry of 4 protomers, presumably two dimers per DNA site.

In a parallel approach to determine the oligomerization state of DNA bound and unbound MtrR, a series of dynamic light scattering (DLS) experiments were done. DLS measures the inherent light scattering of a macromolecule, which fluctuates due to the Brownian motion of the macromolecule, as a function of time. From these measurements the translational diffusion coefficient ( $D_T$ ) of the macromolecule can be calculated, which in turn allows the determination of the hydrodynamic radius of the average scattering particle ( $R_H$ ) via the equation,

$$D_T = kT/6\pi\eta R_H,$$

where  $k$  is the Boltzmann constant,  $T$  is the temperature in Kelvin and  $\eta$  is the solvent viscosity (Garcia de la Torre and Bloomfield 1981).  $R_H$  can be used to estimate the molecular weight.

DLS experiments were carried out at 24 °C on MtrR (50  $\mu$ L of a 0.4 mM dimer solution in PSB) and revealed a MW of 40 kDa  $\pm$  15 kDa, which is consistent with an MtrR dimer. DLS studies on the DNA bound form of MtrR (50  $\mu$ L of a solution containing 0.1 mM dimer MtrR and duplex 27mer) revealed a species with a molecular weight of 110 kDa  $\pm$  10 kDa, which can be explained by the binding of four MtrR protomers ( $4 \cdot 24$  kDa/protomer = 96 kDa) to one 27 bp oligodeoxynucleotide ( $27$  bp  $\cdot$  660 Da/bp = 18 kDa), i.e., 96 kDa + 18 kDa = 114 kDa. The DNA binding data combined with the results from the DLS experiments indicate that two MtrR dimers bind the *mtrCDE* promoter. This DNA binding stoichiometry is the same as that utilized by TetR family member QacR but contrasts with those of family members TetR and EthR, which bind one and four dimers to their respective operators (Hillen and Berens 1994; Engohang-Ndong, Baillat et al. 2004).

To characterize the DNA binding mechanism of MtrR further, the effect of salt concentration on affinity was examined. DNA binding was affected significantly by increasing the NaCl concentration with the  $K_d$  increasing over 100 fold (from 0.9 nM to 99.0 nM) by simply doubling the NaCl concentration from 100 mM, the more physiologically relevant concentration, to 200 mM (Figure 1B). By contrast, when the same experiment was carried out on QacR binding to IR1, only a four fold effect was observed, whereby the  $K_d$  in 100 mM NaCl was 5.7 nM and in 200 mM NaCl, 22.5 nM (Figure 1C). In an attempt to provide a molecular understanding to the different salt effects of MtrR and QacR, both of which bind two dimers to their cognate DNA,

the sequences of the MtrR and QacR DNA binding domains were aligned and an analysis of potential protein-DNA ionic interactions was done by homology modelling (data not shown). QacR engages in only three side chain-phosphate backbone ionic interactions per subunit, thereby providing a reasonable chemical rationale for the modest effect of higher salt on binding affinity. If MtrR were to bind its cognate DNA site in a manner similar to QacR, only those interactions made by QacR would again be made by MtrR as the sequence alignment and homology modelling do not reveal any potential additional ionic interactions within the established QacR DNA binding domain. However, whilst the DNA binding domain of QacR begins with its most N-terminal residue, MtrR has eight additional residues (1-MRKTKEA-8) that are N-terminal to the beginning of the consensus TetR family DNA binding domain. Moreover, MtrR residue 10 is a lysine and the corresponding residue in QacR is an asparagine. Thus, five of the first ten residues of MtrR are basic and not present in QacR and their presence suggests that one or more of these basic residues engages in electrostatic interaction with the *mtrCDE* DNA.

**Secondary structure determination.** As a member of the TetR family, MtrR is expected to be predominantly  $\alpha$  helical. To quantify the secondary structure content of MtrR, circular dichroism (CD) studies were done on MtrR and for further comparison on QacR. Because the crystal structures of apo QacR and a QacR-DNA complex are known (Schumacher, Miller et al. 2001; Schumacher, Miller et al. 2002), quantification of its secondary structure content by CD in these states provides a good idea of the accuracy and precision of this approach in determining the helicity of MtrR. Spectra of the apo and DNA-bound proteins were taken in order to determine whether or not DNA binding significantly alters the secondary structure content of

MtrR. Spectra of the apo proteins and their DNA-bound complexes, in PSB, were taken from 190 nm to 300 nm in a 0.4 mL cell at 25 °C and analyzed for secondary structure content with the deconvolution program K2D (Andrade, Chacon et al. 1993). The concentration of MtrR and QacR in all spectral measurements was 4.0  $\mu$ M dimer and 5.6  $\mu$ M dimer, respectively. These assured stoichiometric DNA binding under the buffer conditions employed in the experiment.

The analysis of the CD spectra of MtrR reveals a helical content of 38% that does not increase or decrease significantly upon DNA binding (Figure 3A, Table 1). This helical content of MtrR is significantly lower than that observed for QacR in solution (~60%) (Figure 3B), which is underestimated by ~20% when compared to the crystal structure (~75%). In addition, QacR is known to lose helicity upon DNA binding according to the crystal structure, a result that is not evident in the CD spectra of the QacR-DNA complex (Figure 3B, Table 1). MtrR has an unanticipated high random coil content (40-45%) and  $\beta$  sheet structure (~18%) that do not appear to change significantly upon DNA binding. Perhaps unstructured regions of MtrR play a role in the binding of small molecule inducers and undergo coil to helix transitions after binding these effectors. Indeed, induction of  $\alpha$  helicity has been observed in the drug-binding domain of thiostrepton-binding transcription regulator TipAS upon binding drug, as well as in the multidrug binding domain of QacR upon multidrug binding (Kahmann, Sass et al. 2003). Contributing in part to the lower than expected helicity and higher random coil content and apparent  $\beta$  structure of MtrR might be its inherent instability, as this protein loses activity over a period of days when stored at 4 °C (data not shown). Therefore, all CD spectra of MtrR were collected within one hour of its purification. Regardless, the finding of  $\beta$  sheet and significant random coil structure in MtrR makes this TetR regulator unusual, as the three dimensional

structures of all other TetR family members, including TetR (Orth, Schnappinger et al. 2000), QacR (Schumacher, Miller et al. 2001; Schumacher, Miller et al. 2002), CprB (Natsume, Ohnishi et al. 2004) and EthR (Dover, Corsino et al. 2004; Frenois, Engohang-Ndong et al. 2004) are essentially all helical.

The data presented here provide a biochemical characterization of MtrR binding to the *mtrCDE* promoter and an assessment of its solution-state secondary structure in the presence and absence of cognate DNA. Unexpectedly, MtrR contains a large amount of random coil and  $\beta$  sheet seemingly beyond the error associated with our CD experiments. The latter secondary structure never has been observed in a TetR family member and confirmation of its existence will require additional structural studies. The differences between MtrR and QacR, TetR and EthR, underscore the wide variety of DNA binding mechanisms of the TetR family. The crystallizations and x-ray structure determinations of MtrR and its DNA complex will be necessary to understand fully the DNA binding mechanism of this TetR family member and are underway.

This work was supported by grants from the American Heart Association (0310050Z to K.M.H.) and the National Institutes of Health (AI21150 to W.M.S.) and (AI48953 to R.G. B.). W.M.S. is the recipient of a Senior Research Career Scientist Award from the VA Medical Research Service.

Table 1. Secondary structure content (%) of MtrR and QacR

	$\alpha$ Helix	$\beta$ Structure	Random Coil
MtrR	37.7	18.3	45.1
MtrR/DNA	38.8	19.4	40.6
QacR	58.3	7.9	33.5
QacR/DNA	60.0	8.0	32.1

## Figure Legends

**Figure 1.** Binding isotherms of MtrR and QacR. **A.** MtrR binding to its 22mer (red curve) and 27mer (black curve) cognate oligodeoxynucleotides. **B.** MtrR binding to the 27mer cognate oligodeoxynucleotide in the presence of 100 mM NaCl (black curve) and 200 mM NaCl (blue curve). **C.** QacR binding to its 28 base pair high affinity DNA binding site (IR1) in the presence of 100 NaCl (orange curve) and 200 mM NaCl (violet curve). The sequence of the one of the IR1 strands is 5'-CTTATAGACCGATCGCACGGTCTATAAG-3'. The binding data displayed in each panel have been normalized to the calculated binding maximum millipolarization of each curve.

**Figure 2.** Determination of the stoichiometry of MtrR-DNA binding. Note the inflection point at an MtrR monomer concentration of 80 nM (black arrow) indicating the shift from high to low affinity binding (indicated with black lines).

**Figure 3.** Circular dichroism spectra of MtrR and QacR. **A.** MtrR in its apo (blue spectra) and DNA-bound (red spectra) forms. **B.** QacR in its apo (orange spectra) and IR1-bound (violet spectra).

Figure 1

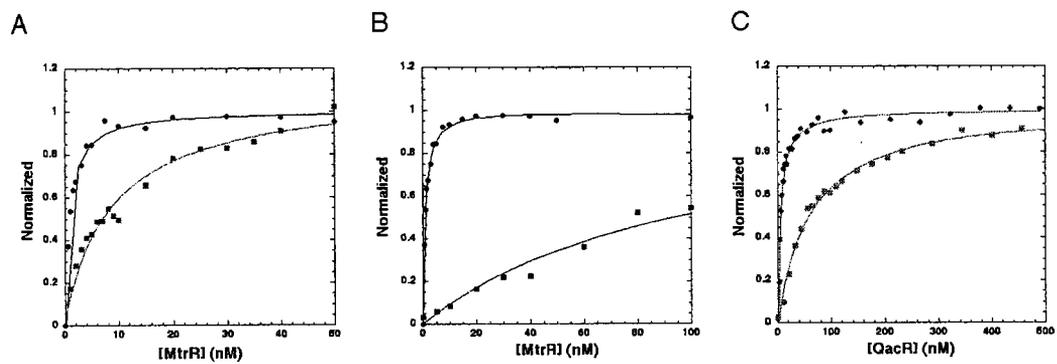


Figure 2

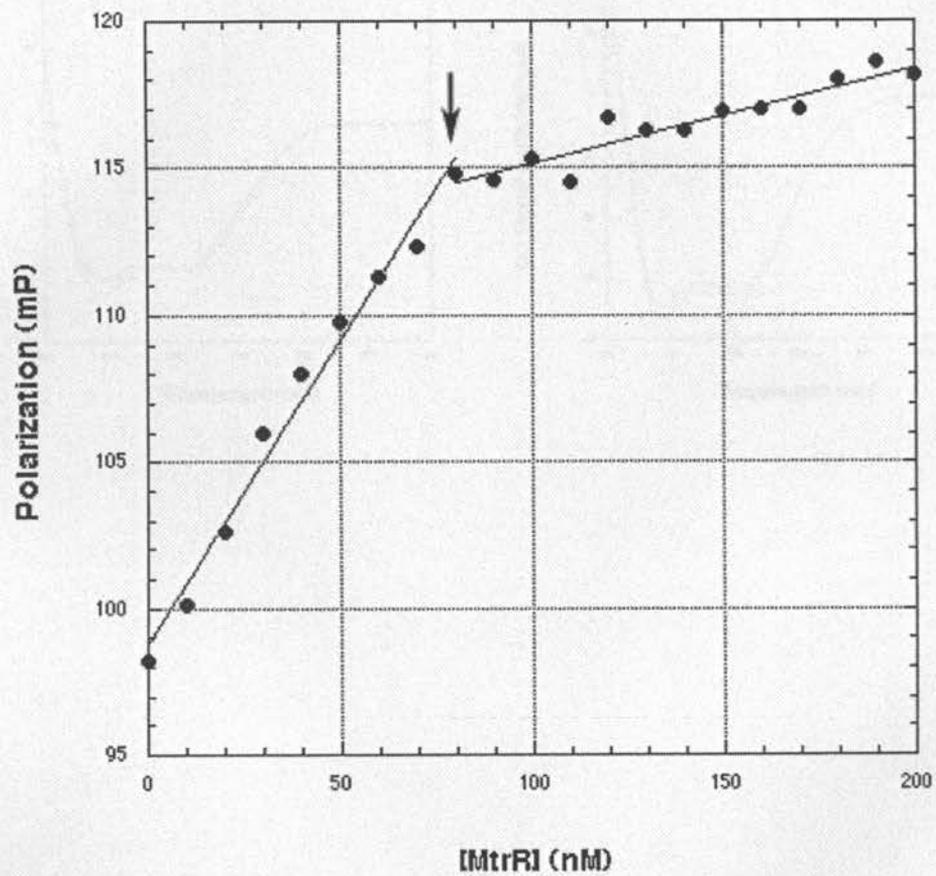
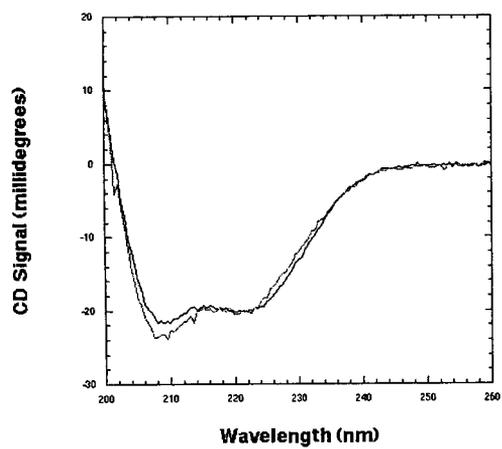
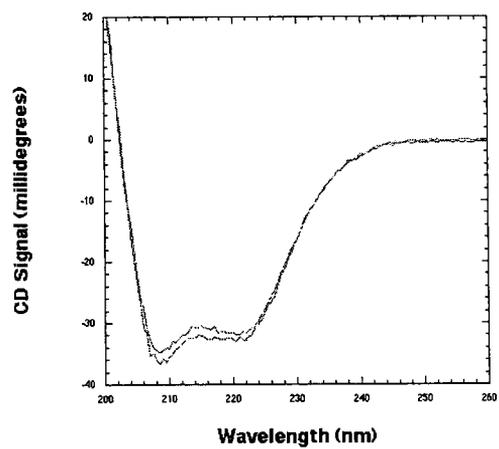


Figure 3

A



B



**Chapter 5: The Role of a Basic N-Terminal Extension of the *Neisseria gonorrhoeae* Multiple Transferable Resistance Repressor, MtrR, in DNA Binding Affinity and Specificity**

Katherine M. Hoffmann<sup>1</sup>, Muthiah Kumaraswami<sup>4</sup>, Jaqueline Pickel<sup>2</sup>, William M. Shafer<sup>2,3</sup>  
and Richard G. Brennan<sup>4</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, Oregon

<sup>2</sup>Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, Georgia

<sup>3</sup>Laboratories of Microbial Pathogenesis, VA Medical Research Service, Veterans Affairs Medical Center, Decatur, Georgia

<sup>4</sup>Department of Biochemistry and Molecular Biology, University of Texas, M.D. Anderson Cancer Center, Houston, Texas

## Abstract

*Neisseria gonorrhoeae* (GC)<sup>1</sup> is a Gram negative bacterium and the causative agent of gonorrhea. Multidrug resistant strains of GC have emerged due in part to overproduction of the multidrug efflux pump, MtrCDE. Expression of the *mtrCDE* genes is repressed by MtrR, which is a member of the TetR family of transcription regulators. Here, we demonstrate that the previously uncharacterized N-terminus of MtrR contains a basic motif, ArgLysXaaLys (RKXX), which is conserved amongst a subclass of TetR proteins that regulate drug and toxin efflux pump genes. This motif interacts with and specifies the A-tract, which is located at one 5' end of a high affinity pseudo direct repeat (mtrDR27) that encompasses the -35 box of the *mtrCDE* promoter. Substitution of these basic residues with alanine or serine or their deletion (D8 or D10) lowers the DNA binding affinity for mtrDR27 by at least 19 fold in 150 mM NaCl. Moreover, whereas wild type MtrR displays 20 fold lower DNA binding affinity when the A-tract is replaced by a G-tract (mtrDR27-G), the D10 protein shows the same, albeit lower, binding affinity for mtrDR27 and mtrDR27-G. Additional mutagenesis and binding studies demonstrate that each basic residue of the RKXX motif contributes equally to DNA affinity and this motif is the major contributor to the lowered affinity of MtrR for cognate DNA in 200 mM NaCl. This newly characterized N-terminal basic motif is expected to play a significant role in the DNA binding affinity and sequence specificity of multiple TetR family members.

## Introduction

Prokaryotic transcriptional regulators are classified into families on the basis of sequence similarity and structural and functional criteria.(Henikoff and Wallace 1988; Aramaki, Yagi et al. 1995; Haft, Loftus et al. 2001; Martinez-Bueno, Molina-Henares et al. 2004; Ramos, Martinez-Bueno et al. 2005) One major group is the TetR family, the members of which are identified through the sequence similarity of their HTH DNA-binding domains (Figure 5.01). Recently, Ramos and colleagues have used a profile matching approach to identify TetR family members and have identified 2,348 nucleic acid and protein sequences that fall into this family.(Ramos, Martinez-Bueno et al. 2005)

TetR family members control genes involved in wide variety of metabolic pathways, responses to environmental stresses, multidrug resistance and pathogenicity(Aramaki, Yagi et al. 1995). Of the 85 TetR family members whose functions have been described, 30, which currently represents the largest functional subgroup, have been identified as regulators of efflux transporters that are involved in antibiotic/multidrug resistance and tolerance to toxic chemicals (Martinez-Bueno, Molina-Henares et al. 2004). One of these proteins is the multiple transferable resistance repressor (MtrR) from *Neisseria gonorrhoeae*. MtrR represses transcription of the *mtrCDE* genes, which encode the MtrCDE multidrug efflux pump (Hagman, Pan et al. 1995; Hagman and Shafer 1995; Lucas 1997; Veal, Yellen et al. 1998; Hoffmann, Williams et al. 2005). Central to this tripartite efflux pump is the MtrD multidrug efflux transporter protein, which is a member of the resistance/nodulation/division (RND) transporter family(Hagman, Lucas et al. 1997). MtrD is linked to MtrE(Delahay, Robertson et al. 1997), the outer membrane

channel protein, by MtrC, the membrane fusion protein. The MtrCDE efflux system recognizes and expels diverse antibacterial hydrophobic agents (HAs) and peptides from the cell into the external milieu (Hagman, Pan et al. 1995). Strains of *N. gonorrhoeae* that exhibit hypersensitivity to HAs can be traced back to mutations in the *mtrCDE* operon (Guymon, Walstad et al. 1978; Hagman, Pan et al. 1995; Veal, Yellen et al. 1998). By contrast, reduced expression of *mtrR*, which is transcribed divergently from the *mtrCDE* genes, results in their increased expression and concomitant resistance to HAs (Hagman and Shafer 1995; Lucas 1997). In addition to its direct involvement in the regulation of *mtrCDE* expression, MtrR is involved either directly or indirectly in the regulation of a second efflux pump, *farAB* (fatty acid resistance). The FarAB efflux system utilizes the MtrE outer membrane channel protein as part of its system and also provides resistance against toxic fatty acids (Delahay, Robertson et al. 1997; Lee and Shafer 1999; Lee, Rouquette-Loughlin et al. 2003).

The *mtrR* gene encodes the 210 amino acid residue MtrR protein, ~23 kDa per subunit. MtrR contains a putative N-terminal helix-turn-helix (HTH) motif and amino acid sequence similarity to multiple members of the TetR family with its strongest resemblance to AcrR with 53% sequence identity and 78% sequence homology between their putative DNA binding domains (Hagman and Shafer 1995; Lucas 1997).

Footprinting experiments have shown that MtrR binds to a 40 base pair (bp) region between the -10 and -35 promoter elements of the *mtrR* gene. This region contains an inverted repeat, although further analysis of the binding site indicates a 31 base pair imperfect direct repeat is the high affinity binding site of MtrR (Lucas 1997). Structure

based sequence alignments of MtrR and TetR (Aramaki, Yagi et al. 1995; Orth, Schnappinger et al. 2000) and QacR (Schumacher, Miller et al. 2002) have revealed the conservation of multiple residues of their major groove binding “recognition” helices, including the Tyr-Trp-His triplet (Tyr-Tyr-His in QacR) and a lysine at the C-terminal end of the recognition helix. This lysine is posited to help lock the HTH motif onto the DNA and is conserved in over 77% of the members of the TetR family (Ramos, Martinez-Bueno et al. 2005). As expected the identities of other residues at positions critical for DNA recognition are less stringently conserved, given the different DNA binding site sequences, but are consistent with the hydrophobic or polar requirement of the position. (Ramos, Martinez-Bueno et al. 2005)

The regulation of the *mtrCDE* genes does not depend upon MtrR alone as the AraC-like protein, MtrA, has been shown to activate this multidrug efflux operon (Rouquette, Harmon et al. 1999). More recently MtrF, which is encoded downstream of *mtrR* and a putative cytoplasmic membrane protein under MtrR regulation, has been implicated in high level detergent resistance in conjunction with MtrCDE (Veal and Shafer 2003; Folster and Shafer 2005). The precise role of MtrR in the regulation of the *farAB* genes, the fatty acid resistance pump proteins, is also unclear. Although some degree of control is evident from knock-out studies and similar inverted repeats can be found in the promotor region of both *farAB* and *mtrCDE* genes, direct regulation has yet to be shown, and the role of the transcriptional regulator FarR is also unclear. Additionally, as work continues to be done on MtrR, exploring it’s mechanism of activation, we hope that at some point a ligand will be conclusively shown to be bound by MtrR, activating it to

allow transcription of the MtrCDE efflux pump proteins, but until then we will focus on the mechanism of repression for biochemical and biophysical studies.

In a previous study we showed that MtrR binds a 27 bp sequence (mtrDR27) upstream of the *mtrCDE* genes as a dimer of dimers, a binding mode which is analogous to that utilized by QacR (Hoffmann, Williams et al. 2005). MtrR binds this site, which contains a pseudo direct repeat with a high affinity (0.9 nM). Moreover, the binding affinity of MtrR for this site displayed a strong salt effect whereby doubling the NaCl concentration from 100 to 200 mM resulted in a 100 fold decrease in binding affinity. By contrast, QacR revealed a far less dramatic salt effect. Comparison of the sequences of the DNA binding domains of MtrR to QacR and TetR revealed that MtrR has an eight residue N-terminal extension beyond the first helix (the positioning helix, (Pabo, Aggarwal et al. 1990; Huffman and Brennan 2002)) of the helix-turn-helix motif that contains three basic residues. These residues, and perhaps an additional lysine that would be residue 2 of the positioning helix, were hypothesized to be the origin of the increased salt sensitivity of DNA binding by MtrR. Here, we test directly the importance of these four basic residues of MtrR (R+7, K+6, K+4 and K2) in DNA binding affinity and specificity and their role in the effect of salt concentration on DNA binding by their site directed substitution and a series of fluorescence based, *in vitro* DNA binding studies. In parallel, we characterized the ability of MtrR to bind to other DNA sequences that have been postulated to have functional relevance *in vivo* (Hagman and Shafer 1995; Lucas 1997; Lee, Rouquette-Loughlin et al. 2003).

## Materials and Methods

### Alignment of N-terminal basic extensions in a functional subgroup of the TetR family:

Thirty members of the TetR family, which comprise the subgrouping of proteins identified as transcriptional repressors of pumps that efflux multiple drugs or other toxic compounds from the cell, were aligned (Figure 5.01). Their primary structures were obtained from the Swiss-Prot database (Gasteiger E. 2003). The alignment of the sequences of the first 60 residues of these proteins, i.e., the N-terminal extensions as well as the consensus TetR family DNA binding domain, was performed using the program ClustalW version 1.81 (Thompson 1994) and the Kyoto University Bioinformatics Center GenomeNet Server (<http://align.genome.jp/>).

**Cloning, mutagenesis, expression and purification of MtrR.** As described previously, the 633 base pair *mtrR* gene from *N. gonorrhoeae* strain FA19 was cloned into the pET-15b ampicillin-resistant vector, which creates an N-terminal hexahistidine tag for purification by Ni<sup>2+</sup> affinity column chromatography (Gasteiger E. 2003). Mutations to codons at positions +7, +6, +4 and 2 (figure 5.01) were created using the Quickchange method with primers designed with significant overhangs to account for GC content (Zheng, Baumann et al. 2004). After mutagenesis both strands were sequenced to ensure only the desired substitution was made. Four single site alanine substitutions, R+7A, K+6A, K+4A, and K2A, and four single site serine substitutions, R+7S, K+6S, K+4S and K2S, were made. Moreover, two quadruple mutants, R+7A/K+6A/K+4A/K2A and R+7S/K+6S/K+4S/K2S, were constructed as were two deletion mutants, in which either

the first 8, D8, or first 10, D10, N-terminal residues were removed. The resulting mutated plasmids were transformed into Rosetta-gami B(DE3)pLysS cells. Each mutant *mtrR* gene was overexpressed and their encoded proteins were purified using Ni<sup>2+</sup>-NTA affinity column chromatography as described previously.(Gasteiger E. 2003)

**Determination of the DNA Binding Affinity.** A fluorescence polarization/anisotropy based assay was used to determine the DNA binding affinity of MtrR for a variety of sequences located in either the *mtrCDE* or *farAB* promoters. These oligodeoxynucleotides were fluoresceinated on the 5' end of one of the complementary strands and ranged in length from 19 to 27 base pairs. The oligodeoxynucleotides encompassed the pseudo direct repeat (DR) that in previous footprinting studies was protected by MtrR, a pseudo inverted repeat (IR) site just upstream of the DR and postulated to be an MtrR binding site, or the site of the *farAB* promoter, which was implicated as a potential binding site in previous sequence homology experiments(Lee, Rouquette-Loughlin et al. 2003). The respective sequences of one strand of each are:  
(mtrDR27) 5'-TTTTTATCCGTGCAATCGTGTATGTAT;  
(mtrDR22) 5'-ATCCGTGCAATCGTGTATGTAT;  
(mtrDR-G) 5'-CCCCATCCGTGCAATCGTGTATGTAT;  
(mtrIR) 5'-GATAAAAAGACTTTTTATG; and  
(faR21) 5'-GGATTAAAATATAACTATTA with their described repeats of the mtrDR and mtrIR in bold and those bases of faR21, which match mtrIR. The standard DNA binding buffer used in these studies was 20 mM Tris-HCl pH 7.4 , 0.1 nM fluoresceinated DNA (or higher, provided  $K_d \geq 10$  fold the DNA concentration) and 1  $\mu$ g of poly d(IC),

as nonspecific DNA. The NaCl concentration was either 100, 150 or 200 mM. MtrR was titrated into the binding mixture to final concentrations that were at least 10 fold above the  $K_d$  where possible, at 25 °C, as described previously (Hoffmann, Williams et al. 2005). Briefly, the sample, containing fluorophore, was excited at 490 nm and the fluorescence polarization emission (in millipolarization units) was detected at 530 nm. The titration data were plotted as the increase in millipolarization with binding against the protein concentration. The equilibrium binding constant is determined from these data using the equation,  $P = \frac{(P_{\text{bound}} - P_{\text{free}})[\text{protein}]}{K_d + [\text{protein}]} + P_{\text{free}}$ , where  $P$  is the polarization measured at a given total protein concentration,  $P_{\text{free}}$  is the initial polarization of free fluorescein-labelled DNA, and  $P_{\text{bound}}$  is the maximum polarization of specifically bound DNA.  $[\text{Protein}]_{\text{free}} = [\text{Protein}]_{\text{total}}$  is assumed because the concentration of fluorescein-labelled DNA is 10 fold below the  $K_d$ . The generated hyperbolic curves are fit by nonlinear least squares regression analysis, assuming a bimolecular model such that the  $K_d$  values represent the protein concentration at half maximal ligand binding. Each binding isotherm was plotted using the graphing program Kaleidograph (Lundblad, Laurance et al. 1996).

**Results and Discussion: An A-tract increases MtrR binding affinity** - In a recent study we determined that MtrR binds a 27 base pair pseudo direct repeat (mtrDR27) with high affinity ( $K_d = 0.9$  nM, (Lundblad, Laurance et al. 1996)). mtrDR27 is located in the *mtrCDE* promoter between the -35 and -10 boxes, including the entire -35 box and was first identified by Lucas *et al.*, (9). Removal of the flanking 5'-A-tract representing the -35 box from this fragment results in a 22 base pair core that contains the pseudo direct repeats. MtrR binds this DNA sequence with a nine fold lower affinity ( $K_d = 7.8$  nM ((Lundblad, Laurance et al. 1996)). To test the importance of the sequence identity of the A-tract, we determined the binding affinity of a 27 base pair oligodeoxynucleotide in which the A-tract was replaced by a G-tract (mtrDR27-G). MtrR binds mtrDR27-G with twenty-fold lower affinity than mtrDR27 (Fig. 5.02, Table 5.01). This clearly points out the importance of the 5'-A-tract in binding affinity and specificity, and suggests a mechanism of repression, whereby MtrR directly blocks the binding of transcriptional machinery.

*Inverted Repeats are low affinity binding sites* - In order to test the ability of MtrR to bind other sequences that have been implicated as potentially functionally relevant, the binding affinity of MtrR for the mtrIR and the farR (faR21) DNA sites were determined. These revealed no specific binding up to micromolar concentrations of protein (Fig. 5.02). The poor binding of MtrR to the faR21 sequence further suggests that MtrR regulates the *farR* gene either indirectly or through binding to a different promoter site (Lee and Shafer 1999; Lee, Rouquette-Loughlin et al. 2003).

*The ArgLysXaaLys (RKXR) motif is conserved and functionally important* - The alignment of the N-terminus of MtrR against those of TetR, QacR and 27 members of a functionally related subgroup of the TetR superfamily which regulate toxin and drug efflux pumps, reveals that TetR and QacR are the exceptions in their lack a significant N-terminal basic extension (Fig. 5.01). Nearly half of this subfamily has an eight to eleven residue basic extension. Intriguingly, the pattern and location (residues +7, +6, and +4, see Fig. 5.01) of the basic residues of MtrR within this region are well conserved both within the subset of proteins which have an extension similar in length to MtrR, as well as across those subgroup members that contain longer extensions. More specifically, in the 15 members of the subfamily with extensions ranging from 8 to 11 residues, an arginine or lysine is found at position +7 in 12 proteins, an arginine or lysine at position +6 in 13 proteins and an arginine, lysine or histidine at position +4 in 11 proteins. Even in those proteins with longer extensions basic residues are found frequently in the positions corresponding to +7, (in 8 of 10 proteins) +6, (5 of 10 proteins) and +4 (5 of 10 proteins). Overall, the +7, +6, +4 basic residues are conserved in 80%, 72% and 64% of the members of the subfamily with N-terminal extensions, respectively (Fig. 5.01).

Previous to this alignment, we postulated that the first eight amino acid residues of MtrR contributed to the higher binding affinity of this regulator for cognate DNA as well as the observed salt effect on binding, in which the affinity dropped over 100 fold as the NaCl concentration was raised from 100 to 200 mM (Lundblad, Laurance et al. 1996). To test this, MtrR deletion mutants were created, in which the first eight ( $\Delta 8$ ) or ten ( $\Delta 10$ ) residues were removed from the N-terminus, and their binding affinities for the mtrDR27

oligodeoxynucleotide were determined in 100 mM and 200 mM NaCl. No significant differences in the binding affinities of the  $\Delta 8$  and  $\Delta 10$  mutants for this 27mer were observed in 100 mM NaCl ( $K_d\Delta 8 = 39$  nM and  $K_d\Delta 10 = 87$  nM) (Fig. 5.03a). However, comparison of the binding affinity of wild type MtrR for this DNA in 100 mM NaCl ( $K_d = 0.9$  nM) reveals 40 and 80 fold lower affinity for the  $\Delta 8$  and  $\Delta 10$  mutants, respectively (Figs. 5.02, 5.03a). As noted previously, the binding of wild type MtrR to cognate DNA displayed a strong electrostatic component as demonstrated by a 100 fold decrease in binding affinity upon a 2 fold increase in NaCl concentration ( $K_d = 99$  nM) (Lundblad, Laurance et al. 1996). The deletion mutants show a similar but decreased salt effect in 200 mM NaCl with binding constants of 2.7 mM for  $\Delta 8$  (~69 fold lower affinity in higher salt) and 2.0 mM for  $\Delta 10$  (~23 fold lower affinity in higher salt).

In order to explore the functional relevance of the N-terminal extension further, additional binding studies were carried out using the deletion mutants and an oligodeoxynucleotide in which the A-tract was removed resulting in a 22mer (mtrDR22) containing the direct repeat but not the -35 box, and the 27mer, mtrDR27-G with the -35 box sequence substituted. In 100 mM NaCl, the  $\Delta 10$  mutant binds these oligodeoxynucleotides with affinities very similar to that of mtrDR27, whereby the binding constants are less than two fold different and are 87 nM (mtrDR27), 158 nM (mtrDR22) and 95 nM (mtrDR27-G) (Fig. 5.03b, Table 5.01). This is in sharp contrast to the binding properties of wild type MtrR, which displays a 20 fold decreased affinity for mtrDR27-G and a 9 fold decrease for mtrDR22 in 100 mM NaCl (Table 5.01). That the N-terminal extension is the origin of the salt effect on binding affinity is apparent from

the comparison of the binding constants of the  $\Delta 10$  mutant for mtrDR27 in 100 mM, 150 mM and 200 mM NaCl. Under these conditions the fold change in affinity of  $\Delta 10$  is only a 17-fold decrease upon doubling the NaCl concentration whereas the wild type protein experiences a 110 fold decrease upon doubling the salt concentration (Table 5.01). The  $\Delta 8$  mutant follows a very similar trend, but due to the significant instability of this protein the binding constant determined with 200 mM NaCl has a large error that makes it too unreliable to draw any strong conclusions. Regardless, the  $\Delta 8$  protein still shows a significant mitigation of the salt effect on DNA binding.

*Quadruple Alanine and Quadruple Serine substitutions-* As noted, the eight residue N-terminal extension beyond helix a1 of MtrR contains three basic residues. In addition the second residue of this helix is basic. The role of each residue in DNA binding affinity was tested by the creation of a series of site directed mutations in which each residue was substituted by either alanine or serine. Furthermore, in order to eliminate any unforeseen complications of the D8 and D10 mutants due to the relocation of the hexahistidine tag and thrombin cleavage site closer to the HTH motif and DNA binding site, we created a quadruple alanine substituted mutant, i.e., Arg+7Ala/Lys+6Ala/Lys+4Ala/Lys2Ala (Quadruple A, Fig. 5.01). The DNA binding affinities of all mutant MtrR proteins were determined and compared to that of the wild type MtrR (Fig. 5.04). For these studies 150 mM NaCl was used in the binding buffer in order to facilitate the comparison of each  $K_d$  (Table 5.01). The Quadruple A protein showed a 19 fold loss in binding affinity for mtrDR27 when compared to the wild type protein (Fig. 5.04a). This finding reinforces the conclusion that the N-terminal extension of MtrR plays a significant role in DNA

binding affinity. However, in absolute terms the binding affinity of the Quadruple A mutant is about four fold better than that observed for the D8 and D10 proteins, indicating that these deletion mutants not only remove residues that contribute to DNA binding affinity but also introduce residues from the affinity tag that interfere somewhat with DNA binding. Single substitution of Lys2 with alanine (Lys2Ala), had the smallest effect on binding with only a 3.7 fold decrease in affinity (Fig. 5.04a). Even this small decrease was somewhat unanticipated as the corresponding glutamine residue in QacR does not interact with DNA (Orth, Schnappinger et al. 2000; Schumacher, Miller et al. 2001; Schumacher, Miller et al. 2002; Schumacher and Brennan 2003; Dover, Corsino et al. 2004; Frenois, Engohang-Ndong et al. 2004). The Lys+6Ala protein displays a binding affinity of 74 nM, a nearly 7 fold increase in  $K_d$  over that of wild type. The Lys+6Ala affinity is the same as those of the Arg+7Ala and Lys+4Ala proteins, which had identical affinities of 66 nM and 65 nM. Given that each of the binding constants of the Arg+7Ala, Lys+6Ala and Lys+4Ala proteins are the same and approximately one third the  $K_d$  of the Quadruple A protein (Table 5.01), we would conclude that no single basic residue of the N-terminal extension is responsible for the salt dependence on DNA binding affinity, but rather each of these residues contributes equally.

To ensure that substitution of residues +7, +6 and +4 and to a lesser extent residue 2, to alanine, singly or *in toto*, did not perturb the structure of the N-terminal extension and hence was the basis for the observed differences in DNA binding affinities, we constructed the analogous series of serine containing mutants, i.e., Arg+7Ser, Lys+6Ser, Lys+4Ser, Lys2Ser and Arg+7Ser/Lys+6Ser/Lys+4Ser/Lys2Ser (Quadruple S) and

measured their binding constants under the same conditions. These serine mutants displayed the same binding trends to *mtrDR27* as their alanine counterparts (Fig. 5.04b, Table 5.01). The slightly lower affinities of the serine substituted proteins for this DNA site is due to their poor stability in the purification buffer. Indeed, storage buffer, which contains a higher ionic strength, stabilizes better both the wild type MtrR and all MtrR site directed mutants. Such stabilization allowed the direct assessment of the effect of raising the salt concentration from 100 mM to 150 mM to 200 mM on the binding affinity of the Quadruple A and Quadruple S proteins. As anticipated, the  $K_d$  rose with increasing salt concentration but by only ~15 fold for Quadruple A. The Quadruple S protein showed less than a two fold decrease in binding affinity.

*Conclusions* - The data presented here confirms that the highly preferred *mtrCDE* DNA binding site of MtrR is the pseudo direct repeat (*mtrDR27*) that includes the -35 box of this promoter and was first identified in footprinting assays nearly ten years ago (9). Furthermore, our studies reveal that MtrR does not bind an upstream inverted repeat (*mtrIR*) in a physiologically relevant manner. Moreover, our binding data, which demonstrate that MtrR has a very low affinity for the *far21* DNA site, indicate that the role of MtrR in the regulation of the *farR* gene in the *farAB* fatty acid efflux pump circuit is indirect or is carried out directly utilizing a different site in the *farR* promoter. Our data do not rule out a role for MtrR in the regulation of the transcription of other regulatory components of this circuit such as an alternate site at FarR or MtrA, but they do underscore that the circuit is more complex than a simple one protein regulatory system.

As a group, the basic residues of the N-terminal extension of MtrR contribute significantly to affinity (compare the  $K_d$  of the wild type MtrR-mtrDR27 complex to those of Quadruple A and Quadruple S-mtrDR27, Table 5.01) and specificity for the -35 box containing cognate DNA (compare the  $K_d$  of MtrR-mtrDR27 to that of the MtrR-mtrDR27-G complex). Moreover, we conclude that the RKXX motif is the major contributor to the salt effect on DNA binding by MtrR with residues Arg+7, Lys+6, and Lys+4 contributing equally to this effect on DNA binding affinity.

Interestingly, our previous CD studies showed that MtrR has significantly more extended structure than the nearly all helical QacR protein as well as the those members of the greater TetR family, the structures of which have been determined (Orth, Schnappinger et al. 2000; Schumacher, Miller et al. 2001; Schumacher, Miller et al. 2002; Schumacher and Brennan 2003; Dover, Corsino et al. 2004; Frenois, Engohang-Ndong et al. 2004). Thus, the N-terminal residues of MtrR likely take a nonhelical structure as only an extended structure could physically reach the upstream A-tract/-35 box. In accord with the idea that this extension is nonhelical, the crystal structure of TetR family member, CprB, reveals the six residue N-terminal extension, MetAlaArgGln,LeuArg (where the underline residues align with the RKXR motif), is not part of helix a1 and disordered in the absence of DNA (Natsume, Ohnishi et al. 2004). Further, the crystal structure of the Ethionamide Repressor (EthR) reveals its twenty-one residue N-terminal extension, which contains an RXXR motif that also aligns with the RKXR motif of MtrR (Fig. 1), is also disordered in the absence of DNA (Orth, Schnappinger et al. 2000; Schumacher,

Miller et al. 2001; Schumacher, Miller et al. 2002; Schumacher and Brennan 2003; Dover, Corsino et al. 2004; Frenois, Engohang-Ndong et al. 2004). Unlike other basic extensions that are involved in DNA binding specificity, e.g., the unstructured six residue basis extension of lambda repressor (Clarke, Beamer et al. 1991), that of MtrR appears to use indirect readout to interact with phosphate backbone, hence the observed salt effects and significantly lower affinity for mtrDR27-G. Perhaps this basic extension straddles the phosphate backbone of the narrowed minor groove of the A-tract, as we have modeled based on the QacR HTH (figure 5.05). Finally and to the best of our knowledge, this is the first characterization of the role of the basic N-terminal extension of any TetR family member in DNA binding. Given that multiple members of the multidrug, toxin, and antibiotic efflux pump gene regulatory subfamily have similar N-terminal extensions, many with conserved basic residues at positions corresponding to +7, +6 and +4 of MtrR (Fig. 5.01), analogous roles of these structures in cognate DNA binding are anticipated.

\*This work was supported by grants from the American Heart Association (0310050Z to K.M.H.) and the National Institutes of Health (AI21150 to W.M.S.) and (AI48953 to R.G. B.). W.M.S. is the recipient of a Senior Research Career Scientist Award from the VA Medical Research Service and R.G.B. is holder of the Robert A. Welch Distinguished University Chair in Chemistry.

<sup>1</sup>GC, *Neisseria gonorrhoeae*, MtrR, Multiple transferable resistance Repressor; HAs, hydrophobic agents; DR, direct repeat; IR, inverted repeat.

<sup>2</sup>The numbering of the MtrR residues that are the foci of this study is based on the sequence alignment of the helix-turn-helix motifs of QacR and MtrR whereby MtrR residues Met<sup>1</sup>, Arg<sup>2</sup>, Lys<sup>3</sup>, Thr<sup>4</sup>, Lys<sup>5</sup>, Thr<sup>6</sup>, Glu<sup>7</sup>, Ala<sup>8</sup>, Leu<sup>9</sup>, and Lys<sup>10</sup> are renumbered to Met+8, Arg+7, Lys+6, Thr+5, Lys+4, Thr+3, Glu+2, Ala+1, Leu1, and Lys2, respectively. This numbering system reflects the addition of residues beyond the positioning helix of the canonical TetR family helix-turn-helix motif.

## Figure Legends

**Table 5.01. Binding constants for all MtrR mutants and all DNA sites reported in this paper.** Fluorescence anisotropy titration binding experiments were performed in 20 mM Tris-HCl pH 7.4, 0.1 nM 5'-fluoresceinated DNA, 1  $\mu$ g of poly d(IC), as nonspecific DNA and concentrations of NaCl as indicated at 25°C. The DNA binding sites indicated encompassed the pseudo direct repeat including the full length site with flanking A-tracts (MtrDR27), the same site with the A-tracts removed (MtrDR22), and replaced by G-tracts for sequence specificity experiments (MtrDR27-G); a pseudo inverted repeat site just upstream of the footprinted site postulated to be an MtrR binding site (MtrIR), or the site of the *farAB* promoter, which was implicated as a potential binding site in previous sequence homology experiments (FaR21). The MtrR variants used include the wild-type protein, deletion mutants of the first 8 and 10 amino acids of MtrR in the N-terminal extension region ( $\Delta$ 8 and  $\Delta$ 10), and site mutants of the basic residues in those first 10 amino acids, specifically R+7, K+6, K+4 and K2 to alanine and serine, as well as the quadruple mutant of these for comparison to the deletion constructs. Dissociation constants ( $K_d$ ) and error are indicated in nanomolar scale.

**Figure 5.01. Aligned sequences of the N-terminal extensions of 30 TetR sub-family members.** These proteins have been identified functionally as regulators of the transcription of multidrug or toxic compound efflux transporters. The sequences of the first 35 residues of their highly conserved TetR DNA binding domains form the basis of the alignment. Consensus sequences are shown at the bottom of the alignment. The position denoted by the number 1 is the first residue of the positioning helix of the Helix-

Turn-Helix domain. Positions of the conserved basic residues of the N-terminal extensions are designated +4, +6 and +7 and marked with blue stars. Sequences that are identical in half or more proteins are shaded black, written in white and homologous sequences, in half or more of the aligned proteins are shaded grey with black letters. The Swiss-prot or gi accession numbers for the protein sequences used are: QacR (P23217, P0A0N3) (Grkovic, Brown et al. 1998; Kuroda, Ohta et al. 2001), PqrA (Q9F147) (Cho, Kim et al. 2003), TetR (P0ACT4) (Unger, Klock et al. 1984), LmrA (O34619) (Kumano, Fujita et al. 2003), UrdK (Q9RP98) (Faust, Hoffmeister et al. 2000), MtrR (P39897) (Pan and Spratt 1994; Hagman and Shafer 1995), IfeR (O68442) (Palumbo, Kado et al. 1998), HydR (gi:15011955) (Farrow, Lyras et al. 2001), AcrR (P34000) (Ma, Cook et al. 1995; Ma, Alberti et al. 1996), EnvR (P31676) (Klein, Henrich et al. 1991), ArpR (Q9KJC4) (Kieboom and de Bont 2001), TtgR (Q9AIU0) (Duque, Segura et al. 2001; Teran, Felipe et al. 2003), BpeR (Q6VV70) (Chan, Tan et al. 2004), SrpR (Q9R9T9) (Kieboom and de Bont 2001), TtgW (Q93PU7) (Rojas, Segura et al. 2003), LanK (Q9ZGB7) (Rebets, Ostash et al. 2003), AmrR (Q9RG61) (Westbrock-Wadman, Sherman et al. 1999), SmeT (Q8KLP4) (Sanchez, Alonso et al. 2002), YdeS (P96676) (Beloin, Ayora et al. 1997), LfrR (gi:61109877) (Li, Zhang et al. 2004), RmrR (Q9KIH5) (Gonzalez-Pasayo and Martinez-Romero 2000), EthR (P96222) (Baulard, Betts et al. 2000), TcmR (P39885) (Guilfoile and Hutchinson 1992; Guilfoile and Hutchinson 1992), Sim16 (Q9AMH9) (Trefzer, Pelzer et al. 2002), AmeR (Q9F8V9) (Peng and Nester 2001), RifQ (O52558) (August, Tang et al. 1998), ActII (Q53901) (Caballero, Malpartida et al. 1991), VarR (Q9AJL5) (Namwat, Lee et al. 2001), Pip (Q9F0Y2) (Folcher, Morris et al. 2001), HemR

(P72185) (Hashimoto, Yamashita et al. 1997). Alignment was done using ClustalW, figure and shading were done using the program Boxshade v3.21

**Figure 5.02. Wild-type MtrR Binding to DNA Sites Proposed in the Literature.** A fluorescence polarization/anisotropy based assay was used to determine the DNA binding affinity of MtrR for 19 to 27 base pairs sequences located in either the *mtrCDE* or *farAB* promoters; the 5'-fluoresceinated oligodeoxynucleotides encompassed the pseudo direct repeat (DR) that in previous footprinting studies was protected by MtrR, a pseudo inverted repeat (IR) site just upstream of the DR and postulated to be an MtrR binding site, or the site of the *farAB* promoter, which was implicated as a potential binding site in previous sequence homology experiments. The DNA binding buffer used in these studies was 20 mM Tris-HCl pH 7.4, 0.1 nM fluoresceinated DNA, 1  $\mu$ g of poly d(IC), as nonspecific DNA and 100 mM NaCl. MtrR was titrated into the binding mixture to final concentrations that were at least 10 fold above the  $K_d$  where possible, at 25 °C. Black squares = binding to the 27 base pair direct repeat site identified by footprinting assay (MtrDR27); green squares = the footprinting site with the flanking A-tracts replaced by G-tracts (MtrDR27-G); yellow circles, binding to the inverted repeat upstream in the *mtrCDE* promoter (MtrIR); purple circles, binding to the homologous inverted repeat in the *farAB* promoter region (FaR21).

**Figure 5.03. Salt Effect in MtrR Deletion Mutants and Specificity for DNA Sequence.**

Fluorescence anisotropy titration binding studies with deletion mutants of the first 8 and first 10 amino acid residues of MtrR in 20 mM Tris-HCl pH 7.4, 0.1 nM fluoresceinated DNA, 1  $\mu$ g of poly d(IC), as nonspecific DNA and 100 mM NaCl, unless otherwise indicated. A: Salt effect

comparison studies of the  $\Delta 8$  (blue) and  $\Delta 10$  (red) mutants in which the NaCl concentration in the binding buffer was varied; binding in 100 mM NaCl conditions are indicated by triangles, binding in 200 mM NaCl indicated by squares. B: Specificity for DNA sequence in the flanking tracts. Binding of the  $\Delta 10$  mutant to the full-length 27-mer direct repeat site (red triangles, MtrDR27 as in A), the same site with the flanking A-tracts removed (MtrDR22, yellow squares) and with the flanking A-tracts replaced by G-tracts (green circles, MtrDR27-G).

**Figure 5.04. Binding Affinity of MtrR Site Mutants To the Full Length Direct**

**Repeat DNA Site.** Fluorescence anisotropy titration binding studies of wild-type (black squares), single site (diamonds), quadruple site (green circles) and  $\Delta 8$  and  $\Delta 10$  deletion mutants (blue and red circles, respectively) in 20 mM Tris-HCl pH 7.4, 0.1 nM fluoresceinated MtrDR27 DNA, 1  $\mu$ g of poly d(IC), as nonspecific DNA and 150 mM NaCl binding conditions. Single site mutants are indicated with the following colors: R+7 is in pink, K+6 is light blue, K+4 is yellow and K2 is orange. A: Single site and quadruple mutations of basic residues to alanine. B: Single site and quadruple mutations of basic residues to serine.

**Fig. 5.05. A model of DNA binding by the RKXXK motif.** The eight residue N-terminal extension of MtrR was modelled onto helix a1 of QacR bound to DNA (1QPZ). The DNA binding domain of one subunit including helices a1 through part of a4 is represented by teal ribbons and labelled. The N-terminal extension of MtrR is shown as a green tube with residues of the RKXR motif presented as thicker sticks and coloured according to atom type whereby carbon is green; nitrogen, blue; oxygen, red; hydrogen,

white. The eight residue extension is shown as taking an elongated structure in order to interact with the phosphate backbone of the minor groove of the A-tract of the *mtrCDE* operator.

Table 5.01.

MtrR protein	DNA site	[NaCl] (mM)	K <sub>d</sub> (nM)	Error (nM)
wildtype	mtrDR27	100	0.9 <sup>a</sup>	0.05
wildtype	mtrDR27	150	10 <sup>a</sup>	1
wildtype	mtrDR27	200	99 <sup>a</sup>	22
wildtype	mtrDR22	100	7.8 <sup>a</sup>	1.0
wildtype	mtrDR27-G	100	18	3
wildtype	faR21	100	NSB <sup>b</sup>	
wildtype	mtrIR	100	NSB <sup>b</sup>	
Δ10	mtrDR27	100	87	12
Δ10	mtrDR27	150	800	200
Δ10	mtrDR27	200	1500	700
Δ10	mtrDR22	100	158	24
Δ10	mtrDR27-G	100	95	19
Δ8	mtrDR27	100	39	6
Δ8	mtrDR27	150	700	160
Δ8	mtrDR27	200	2700	1200
quadruple Ala	mtrDR27	150	190	20
Arg+7Ala	mtrDR27	150	66	14
Lys+6Ala	mtrDR27	150	74	14
Lys+4Ala	mtrDR27	150	65	10
Lys2Ala	mtrDR27	150	37	7
quadruple Ser	mtrDR27	150	440	180
Arg+7Ser	mtrDR27	150	127	20
Lys+6Ser	mtrDR27	150	144	12
Lys+4Ser	mtrDR27	150	131	25
Lys2Ser	mtrDR27	150	45	5

<sup>a</sup> Indicates that these data were previously reported in Hoffmann *et al.*, (2005).

<sup>b</sup> NSB = no specific binding was detected.

```

QacR      1  -----KNLKGYMATTTGEIVKLSSESSKGNLYXHFKTKEML
PqxR      1  -----MAVDR--STATMDEVARAAGLSRATLHRHFAGRDAL
Tetr      1  -----MARLNHRGIDGLTTRKLAQKLGIEQPTLYHVKMKRAL
LmrR      1  -----MSYGDSSRGYYGIGLNQIIKESGAPKOSLYHFGGKEQ
UrdK      1  -----HTTPKSTRGYEKTSLRGIAERLNVTKPAIYHFGTKEAT
HtrR      1  -----HRRKTKTEALKTRGIAARTSLNEIAQAAGVTRGALYHFKMKEDL
IfcR      1  -----HRRKTKTEALKTRGIAARTSLNEIAQAAGVTRGALYHFKMKEDL
HydR      1  -----HRRKTKTEALKTRGIAARTSLNEIAQAAGVTRGALYHFKMKEDL
AcxR      1  -----MARKTKQEAQETRGVSSSTSLGEIAKAGVTRGAIYHFKDKSDL
EnvR      1  -----MAKRTKAEALKTRGVSKITLNDIADANVTRGAIYHFKDKSDL
ArpR      1  -----MVRRTKEEAQETRGVARTTLADIAGVTRGAIYHFKDKSDL
TtgR      1  -----MVRRTKEEAQETRGVARTTLADIAGVTRGAIYHFKDKSDL
BpeR      1  -----HARRTKEEALATRGVSHITSLADIAGVTRGAIYHFKDKSDL
SrpR      1  -----MARKTAAEAETRGVSDATLDQIARKAGVTRGAVYHFKDKSDL
TtgW      1  -----MARKTAAEAETRGVSDATLDQIARKAGVTRGAVYHFKDKSDL
LanX      1  -----HGGTPHVRGANTRGYEKTSHREIAEGLGITKAAALYHFKDKSDL
AmrR      1  -----MARKTKEEESQKTRGVGTIADADADAAGVSRGAVYHFKDKSDL
SmeT      1  -----MARKTKEDTQAETRGVARTTLEHIGARAGYTRGAVYHFKDKSDL
YdeS      1  -----MQSKRGRPRDEGTNGFDVAIVDKIAERAKVSKATIEKWWSSKAAV
LfrR      1  -----MTSPSIESGARERTDHPFTAALGDIAAAAGVGRSTVHRYPERTDL
RmrR      1  -----MVQNHKIEKRFRGRFQVRCDDDTKGYAAAASIAAIAQAGVSTXTLYRFLPTKADL
EthR      1  -----HTTSAASQASLPGRRTARPSQDDREFLADISVDDLAQKAGISRPYFYFYSKEAV
TcmR      1  -----HDSAEIDTPTSTRSTPNGPGLRQRKLRRTRGYEHTIVEQIAEAVEVHPRTFRHFAKSEEV
SimReg2   1  -----MWHNEFVSIWHPPEPAGRRSARSHRTLARGVEAASHRRVAEELGAGTMSLYXYVPTKEDL
AmeR      1  -----HNKTIDQVRKGDRLKSDLFVRRRPRRSAAEETRGFNAVAIADIASALNHSFANVFKHFSKSMAL
RifQ      1  -----MGKRASQPRESVWLSRPRPARKHRAGESELDRGDAKFSHRLLAEELNVTPHSVYHYVANKDDDL
ActII     1  -----MSRSEEGRPKFEEIFVFPWRRPKKAPFRMPLTQGLDALSMRRLAQELKTGHASLYAHVGNRDEL
VarR      1  -----HAAKRASQPRSSVWLTPEFTTRGRRGPERGSGAGSLDRGDAKFTHRVLA TELGVTPHSVYHYVANKDDDL
Pip       1  HMSRGEVRHAXAGREGFRDSVHLSGEGRRGGRRGRQPSGLDRGLTGF SHHRLAEELNVTPHSVYHYVDTKDDQL
HemR      1  HDA SDHNE YF SGL VFGHG GVVADADGQG TIFRETRRRGPALEGYAGLTFEAVAEERHHTGKAAALYRRRPFKESL
consensus      rk k a trgi tm eia agvtrgalyyhf k el
                  +7+6 +4 +1 1...
                  * * *

```

Figure 5.02:

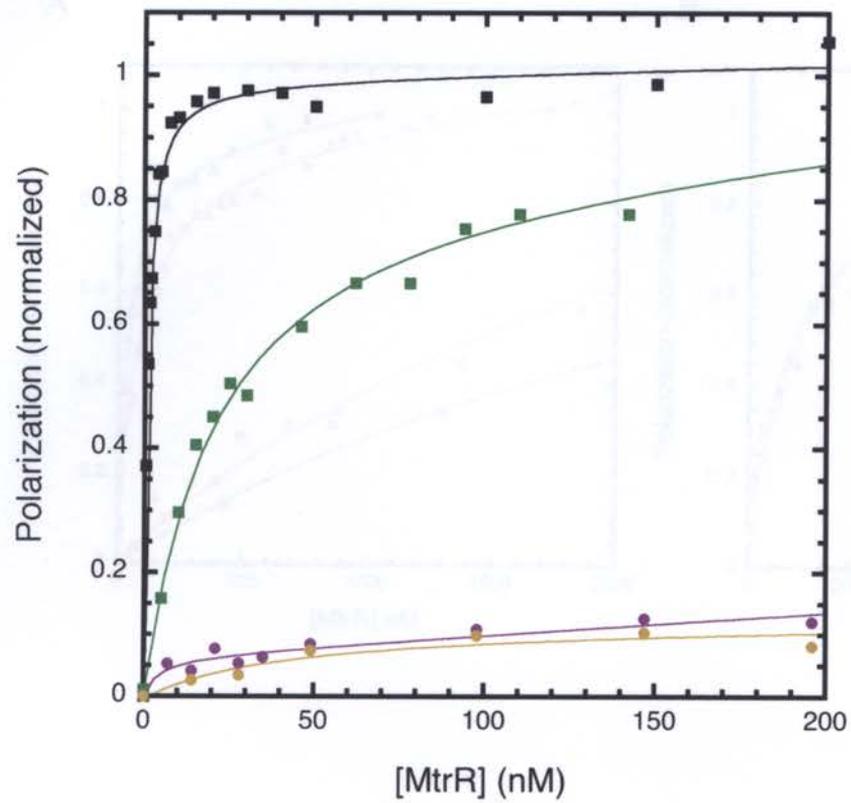


Figure 5.03:

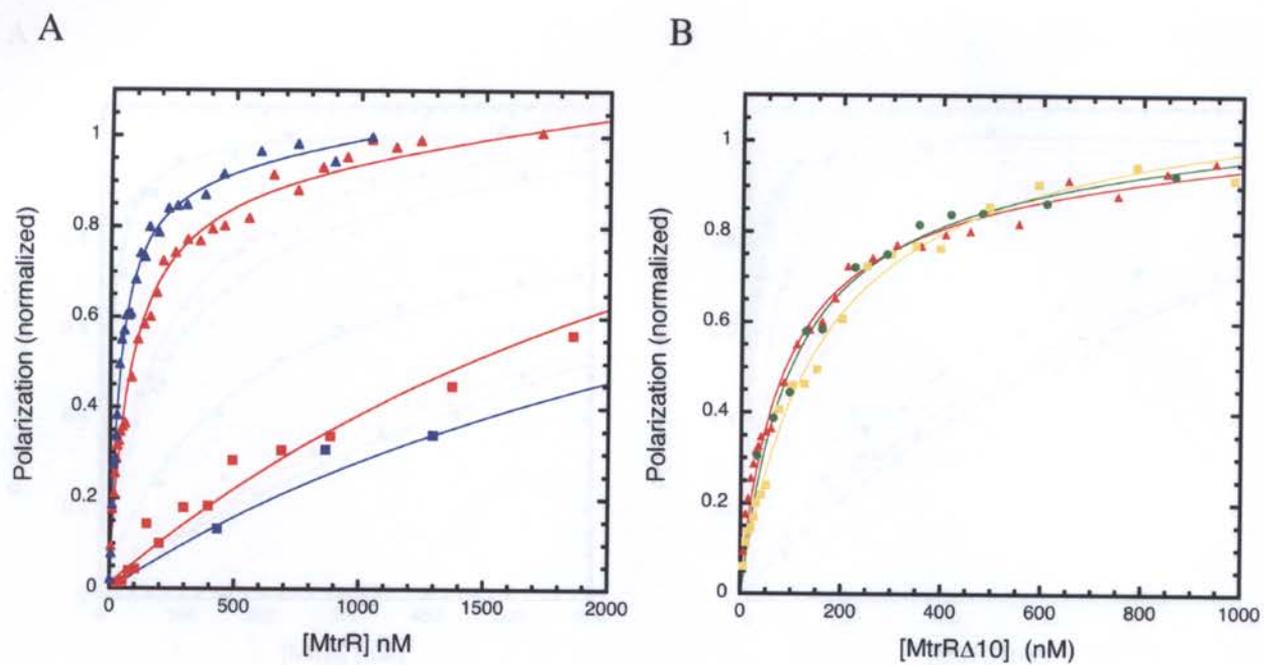
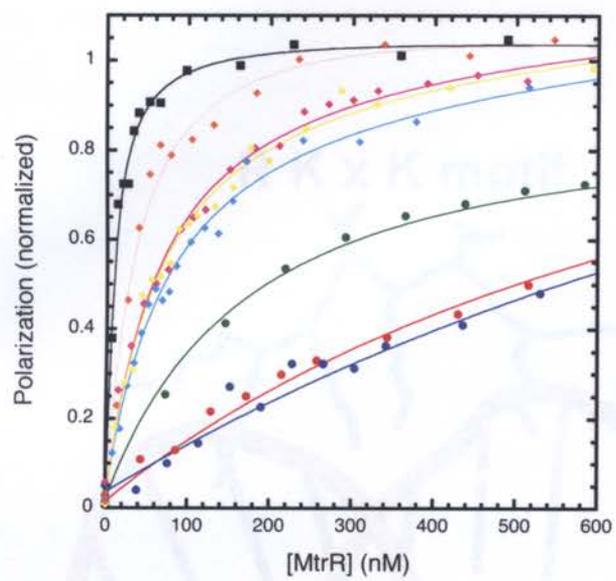


Figure 5.04:

A



B

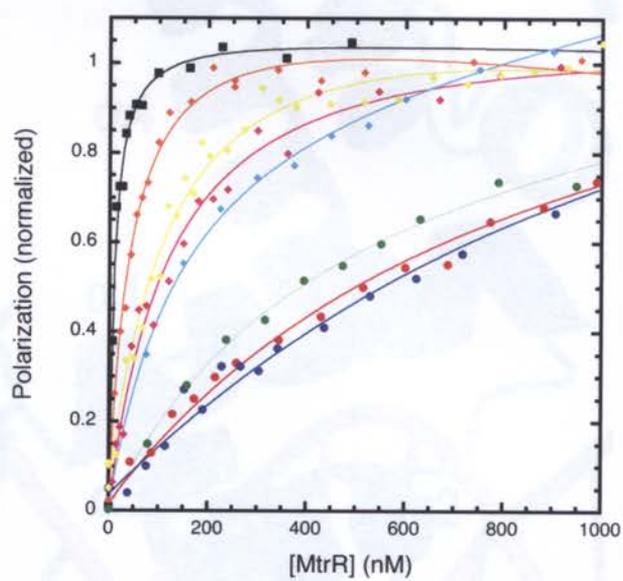
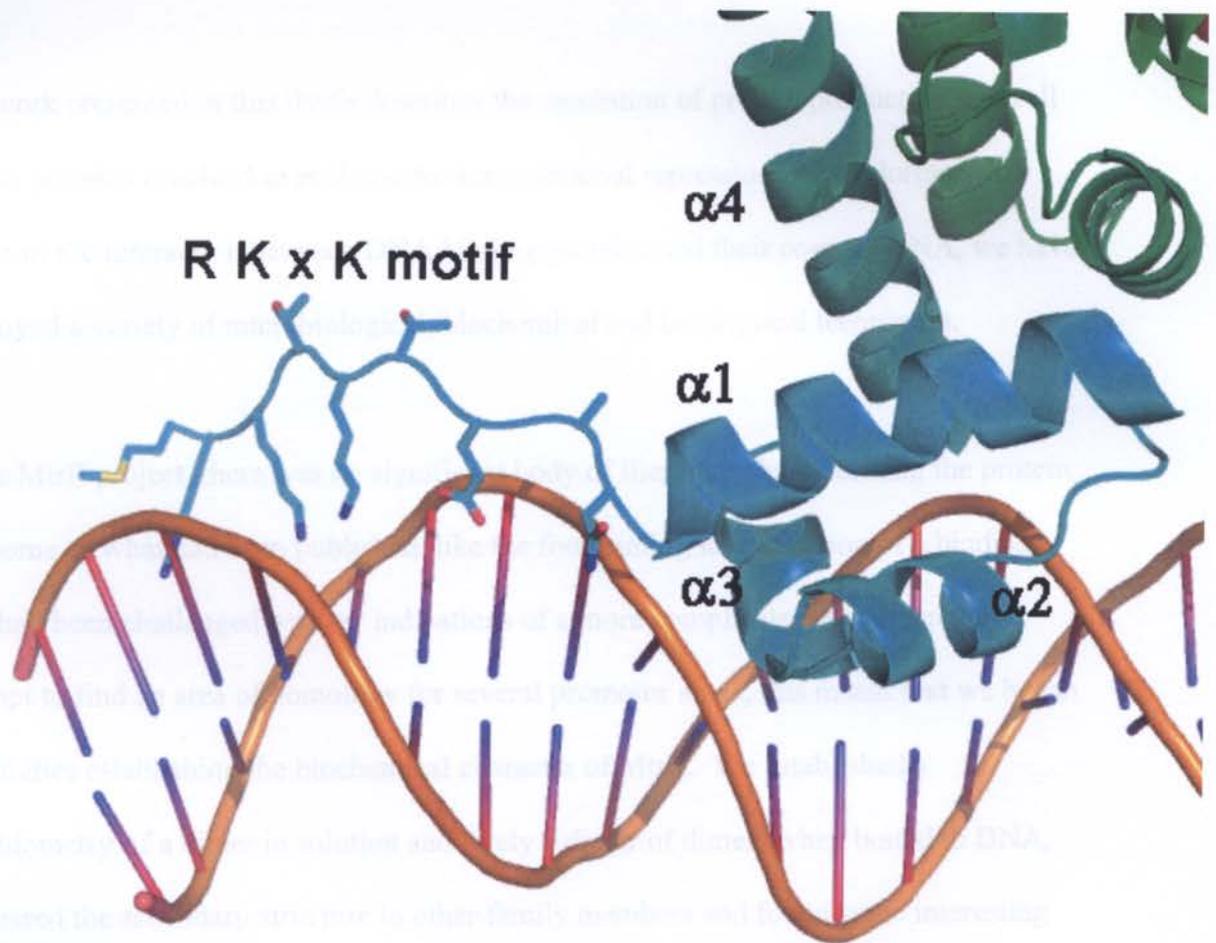


Figure 5. Secondary and Discussion



A lack of studies on Mtk itself does not mean a dearth of information on related proteins, however, and in comparing and contrasting Mtk with other Tdk family members we became very curious about the role of the 3 residues (4 amino acid residues) that had not been seen in any of the crystal structures solved in the family, but which seemed to be highly conserved within the functional sub-family. Metagenomics studies allowed us to identify three basic residues (we'll return to) that have a significant impact

## Chapter 6: Summary and Discussion

The work presented in this thesis describes the regulation of protein production in a cell by two proteins involved in prokaryotic transcriptional repression. In exploring the nature of the interaction between DNA binding proteins and their cognate DNA, we have employed a variety of microbiological, biochemical and biophysical techniques.

In the MtrR project, there was no significant body of literature characterizing the protein, and some of what had been published, like the footprinting identification of a binding site, had been challenged by later indications of a more complicated circuit and an attempt to find an area of homology for several promotor areas; this meant that we began our studies establishing the biochemical character of MtrR. We established a stoichiometry of a dimer in solution and likely a dimer of dimers when bound to DNA, compared the secondary structure to other family members and found some interesting potential differences, confirming the footprinted site and ruled out several others that had been suggested by homology.

A lack of studies on MtrR itself does not mean a dearth of information on related proteins, however, and in comparing and contrasting MtrR with other TetR family members we became very curious about the role of the 8 residue N-terminal extension that hadn't been seen in any of the crystal structures solved for the family, but which seemed to be highly conserved within the functional sub-family. Mutagenesis studies allowed us to identify three basic residues (well conserved) that have a significant impact

on binding if they're removed. We found, additionally, that mutating all three abrogated the strong salt effect we'd established for MtrR (in contrast to QacR, another family member, which doesn't have the N-terminal extension) and also removed MtrR's ability to discern binding site length or the identity of the residues flanking the pseudo-direct repeat. All of this points to the potential importance of the N-terminal extension and indicates a role in interacting with the flanking adenine-rich tracts in the binding site, one of which is a true A-tract, and one of which includes the -10, TATAA box. Whether the sequence of DNA bases is going to influence the binding of MtrR and other family members to their cognate sites through direct or indirect means remains for the crystallography model to reveal.

Adenine tracts are also relevant to the PurR project because the tracts flanking the central CpG kink were so highly conserved in the operon. Previous work mutating the adenine 8 position as well as the K55 protein residue had not revealed dramatic differences in global structure, but mutating one of the center adenines in the tracts did. We found that the only other base naturally occurring in the operon in a single half site, cytosine, still allowed PurR to make wild-type contacts at K55, along the backbone and at the distal HTH-major groove interface with only a mild energetic compensation. Thymine, by contrast, lost some of the contacts to K55, and guanine in that position introduced a barrier to proper structure in the complex and only achieved a 40° bend where the wild type structure displays a 49° kink.

Further attempts to clarify the detailed structure around the substituted base led us to make the mutation in both half-sites to compensate for statistical disorder, an artifact of crystallization in our system. Substituting within the adenine tract in both of the half-sites more than doubled the effect that we saw, letting both the 2Cyt7 and 2Thy7 mutants display global change in DNA structure, and a strain on the hydrogen bonds established in the wild type structure as important for recognizing cognate DNA; indeed, the 2Thy7 structure was unable to make the connections in a few instances, although the overall structure was still close enough to wild-type to bind stably and crystallize.

The PurR indirect readout structures, beyond being interesting relative to the LacI family and theories as to the structural details of association and dissociation, is also unique as a study that demonstrates global changes in DNA distal from a site of indirect readout. The MtrR studies, hopefully to be supported with crystal structures eventually, also has broader implications than circuit of HA resistance in *Neisseria*; the homology that the N-terminal extension shares with the functional subgroup in the TetR family may inform a number of other proteins involved in multidrug efflux, particularly as to how transcriptional regulators identify, discriminate and bind cognate promoters with high affinity.

- (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-3.
- Andrade, M. A., P. Chacon, et al. (1993). "Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network." Protein Eng **6**(4): 383-90.
- Aramaki, H., N. Yagi, et al. (1995). "Residues important for the function of a multihelical DNA binding domain in the new transcription factor family of Cam and Tet repressors." Protein Eng **8**(12): 1259-66.
- Arvidson, D. N., F. Lu, et al. (1998). "The structure of PurR mutant L54M shows an alternative route to DNA kinking." Nat Struct Biol **5**(6): 436-41.
- August, P. R., L. Tang, et al. (1998). "Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699." Chem Biol **5**(2): 69-79.
- Barry, J. K. and K. S. Matthews (1999). "Thermodynamic analysis of unfolding and dissociation in lactose repressor protein." Biochemistry **38**(20): 6520-8.
- Baulard, A. R., J. C. Betts, et al. (2000). "Activation of the pro-drug ethionamide is regulated in mycobacteria." J Biol Chem **275**(36): 28326-31.
- Bell, C. E. and M. Lewis (2000). "A closer view of the conformation of the Lac repressor bound to operator." Nat Struct Biol **7**(3): 209-14.
- Beloin, C., S. Ayora, et al. (1997). "Characterization of an lrp-like (lrpC) gene from *Bacillus subtilis*." Mol Gen Genet **256**(1): 63-71.
- Blundell, T. a. J., LN (1976). Protein Crystallography. San Diego, California, Academic Press, Inc.
- Bosch, D., M. Campillo, et al. (2003). "Binding of proteins to the minor groove of DNA: what are the structural and energetic determinants for kinking a basepair step?" J Comput Chem **24**(6): 682-91.
- Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-21.
- Caballero, J. L., F. Malpartida, et al. (1991). "Transcriptional organization and regulation of an antibiotic export complex in the producing *Streptomyces* culture." Mol Gen Genet **228**(3): 372-80.
- Cantor, C. a. S., P. (1980). Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function. San Francisco, W. H. Freeman.
- Chan, H. W., J. B. Dodgson, et al. (1977). "Influence of DNA structure on the lactose operator-repressor interaction." Biochemistry **16**(11): 2356-66.
- Chan, Y. Y., T. M. Tan, et al. (2004). "BpeAB-OprB, a multidrug efflux pump in *Burkholderia pseudomallei*." Antimicrob Agents Chemother **48**(4): 1128-35.
- Chen, J., R. Surendran, et al. (1994). "Construction of a dimeric repressor: dissection of subunit interfaces in Lac repressor." Biochemistry **33**(5): 1234-41.
- Chen, S., A. Gunasekera, et al. (2001). "Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: alteration of DNA binding specificity through alteration of DNA kinking." J Mol Biol **314**(1): 75-82.

- Chen, S., J. Vojtechovsky, et al. (2001). "Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking." *J Mol Biol* **314**(1): 63-74.
- Cho, Y. H., E. J. Kim, et al. (2003). "The pqrAB operon is responsible for paraquat resistance in *Streptomyces coelicolor*." *J Bacteriol* **185**(23): 6756-63.
- Choi, K. Y. and H. Zalkin (1992). "Structural characterization and corepressor binding of the *Escherichia coli* purine repressor." *J Bacteriol* **174**(19): 6207-14.
- Choi, K. Y. and H. Zalkin (1994). "Role of the purine repressor hinge sequence in repressor function." *J Bacteriol* **176**(6): 1767-72.
- Chuprina, V. P., J. A. Rullmann, et al. (1993). "Structure of the complex of lac repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics." *J Mol Biol* **234**(2): 446-62.
- Clarke, N. D., L. J. Beamer, et al. (1991). "The DNA binding arm of lambda repressor: critical contacts from a flexible region." *Science* **254**(5029): 267-70.
- Delahay, R. M., B. D. Robertson, et al. (1997). "Involvement of the gonococcal MtrE protein in the resistance of *Neisseria gonorrhoeae* to toxic hydrophobic agents." *Microbiology* **143** (Pt 7): 2127-33.
- Dickerson, R. E. (1998). "DNA bending: the prevalence of kinkiness and the virtues of normality." *Nucleic Acids Res* **26**(8): 1906-26.
- Dickerson, R. E. and T. K. Chiu (1997). "Helix bending as a factor in protein/DNA recognition." *Biopolymers* **44**(4): 361-403.
- Dickerson, R. E., D. Goodsell, et al. (1996). "MPD and DNA bending in crystals and in solution." *J Mol Biol* **256**(1): 108-25.
- Dover, L. G., P. E. Corsino, et al. (2004). "Crystal structure of the TetR/CamR family repressor *Mycobacterium tuberculosis* EthR implicated in ethionamide resistance." *J Mol Biol* **340**(5): 1095-105.
- Drenth, J. (1999). *Principles of Protein X-Ray Crystallography*. New York City, Springer-Verlag Inc.
- Duque, E., A. Segura, et al. (2001). "Global and cognate regulators control the expression of the organic solvent efflux pumps TtgABC and TtgDEF of *Pseudomonas putida*." *Mol Microbiol* **39**(4): 1100-6.
- el Hassan, M. A. and C. R. Calladine (1996). "Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA." *J Mol Biol* **259**(1): 95-103.
- el Hassan, M. A. and C. R. Calladine (1996). "Structural mechanics of bent DNA." *Endeavour* **20**(2): 61-7.
- El Hassan, M. A. and C. R. Calladine (1998). "Two distinct modes of protein-induced bending in DNA." *J Mol Biol* **282**(2): 331-43.
- Engohang-Ndong, J., D. Baillat, et al. (2004). "EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator." *Mol Microbiol* **51**(1): 175-88.
- Farrow, K. A., D. Lyras, et al. (2001). "Genomic analysis of the erythromycin resistance element Tn5398 from *Clostridium difficile*." *Microbiology* **147**(Pt 10): 2717-28.
- Faust, B., D. Hoffmeister, et al. (2000). "Two new tailoring enzymes, a glycosyltransferase and an oxygenase, involved in biosynthesis of the angucycline

- antibiotic urdamycin A in *Streptomyces fradiae* Tu2717." Microbiology **146** (Pt 1): 147-54.
- Folcher, M., R. P. Morris, et al. (2001). "A transcriptional regulator of a pristinamycin resistance gene in *Streptomyces coelicolor*." J Biol Chem **276**(2): 1479-85.
- Folster, J. P. and W. M. Shafer (2005). "Regulation of *mtrF* expression in *Neisseria gonorrhoeae* and its role in high-level antimicrobial resistance." J Bacteriol **187**(11): 3713-20.
- Frenois, F., J. Engohang-Ndong, et al. (2004). "Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis." Mol Cell **16**(2): 301-7.
- Friedman, A. M., T. O. Fischmann, et al. (1995). "Crystal structure of lac repressor core tetramer and its implications for DNA looping." Science **268**(5218): 1721-7.
- Garcia de la Torre, J. G. and V. A. Bloomfield (1981). "Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications." Q Rev Biophys **14**(1): 81-139.
- Garvie, C. W. and S. E. Phillips (2000). "Direct and indirect readout in mutant Met repressor-operator complexes." Structure **8**(9): 905-14.
- Garvie, C. W. and C. Wolberger (2001). "Recognition of specific DNA sequences." Mol Cell **8**(5): 937-46.
- Gasteiger E., G. A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. (2003). "ExPASy: the proteomics server for in-depth protein knowledge and analysis." Nucleic Acids Res **31**: 3784-3788.
- George H. Stout, L. H. J. (1989). X-Ray Structure Determination: A Practical Guide, Joh Wiley & Sons, Inc.
- Gincel, E., G. Lancelot, et al. (1994). "Comparison of solution structure of free and complexed lac operator by molecular modelling with NMR constraints." Biochimie **76**(2): 141-51.
- Glasfeld, A., A. N. Koehler, et al. (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**(2): 347-61.
- Gonzalez-Pasayo, R. and E. Martinez-Romero (2000). "Multiresistance genes of *Rhizobium etli* CFN42." Mol Plant Microbe Interact **13**(5): 572-7.
- Greenfield, N. and G. D. Fasman (1969). "Computed circular dichroism spectra for the evaluation of protein conformation." Biochemistry **8**(10): 4108-16.
- Grkovic, S., M. H. Brown, et al. (1998). "QacR is a repressor protein that regulates expression of the *Staphylococcus aureus* multidrug efflux pump QacA." J Biol Chem **273**(29): 18665-73.
- Grkovic, S., M. H. Brown, et al. (2001). "The staphylococcal QacR multidrug regulator binds a correctly spaced operator as a pair of dimers." J Bacteriol **183**(24): 7102-9.
- Grkovic, S., M. H. Brown, et al. (2002). "Regulation of bacterial drug export systems." Microbiol Mol Biol Rev **66**(4): 671-701, table of contents.
- Guilfoile, P. G. and C. R. Hutchinson (1992). "Sequence and transcriptional analysis of the *Streptomyces glaucescens* *tcmAR* tetracenomycin C resistance and repressor gene loci." J Bacteriol **174**(11): 3651-8.

- Guilfoile, P. G. and C. R. Hutchinson (1992). "The *Streptomyces glaucescens* TcmR protein represses transcription of the divergently oriented *tcmR* and *tcmA* genes by binding to an intergenic operator region." J Bacteriol **174**(11): 3659-66.
- Guymon, L. F., D. L. Walstad, et al. (1978). "Cell envelope alterations in antibiotic-sensitive and-resistant strains of *Neisseria gonorrhoeae*." J Bacteriol **136**(1): 391-401.
- Haft, D. H., B. J. Loftus, et al. (2001). "TIGRFAMs: a protein family resource for the functional identification of proteins." Nucleic Acids Res **29**(1): 41-3.
- Hagerman, P. J. (1990). "Sequence-directed curvature of DNA." Annu Rev Biochem **59**: 755-81.
- Hagman, K. E., C. E. Lucas, et al. (1997). "The MtrD protein of *Neisseria gonorrhoeae* is a member of the resistance/nodulation/division protein family constituting part of an efflux system." Microbiology **143** (Pt 7): 2117-25.
- Hagman, K. E., W. Pan, et al. (1995). "Resistance of *Neisseria gonorrhoeae* to antimicrobial hydrophobic agents is modulated by the *mtrRCDE* efflux system." Microbiology **141** (Pt 3): 611-22.
- Hagman, K. E. and W. M. Shafer (1995). "Transcriptional control of the *mtr* efflux system of *Neisseria gonorrhoeae*." J Bacteriol **177**(14): 4162-5.
- Hashimoto, Y., M. Yamashita, et al. (1997). "The *Propionibacterium freudenreichii* hemYHBXRL gene cluster, which encodes enzymes and a regulator involved in the biosynthetic pathway from glutamate to protoheme." Appl Microbiol Biotechnol **47**(4): 385-92.
- He, B., K. Y. Choi, et al. (1993). "Regulation of *Escherichia coli* *glnB*, *prsA*, and *speA* by the purine repressor." J Bacteriol **175**(11): 3598-606.
- He, B., A. Shiau, et al. (1990). "Genes of the *Escherichia coli* *pur* regulon are negatively controlled by a repressor-operator interaction." J Bacteriol **172**(8): 4555-62.
- Henikoff, S. and J. C. Wallace (1988). "Detection of protein similarities using nucleotide sequence databases." Nucleic Acids Res **16**(13): 6191-204.
- Hillen, W. and C. Berens (1994). "Mechanisms underlying expression of Tn10 encoded tetracycline resistance." Annu Rev Microbiol **48**: 345-69.
- Hoffmann, K. M., D. Williams, et al. (2005). "Characterization of the multiple transferable resistance repressor, MtrR, from *Neisseria gonorrhoeae*." J Bacteriol **187**(14): 5008-12.
- Huffman, J. L. and R. G. Brennan (2002). "Prokaryotic transcription regulators: more than just the helix-turn-helix motif." Curr Opin Struct Biol **12**(1): 98-106.
- Huffman, J. L., F. Lu, et al. (2002). "Role of residue 147 in the gene regulatory function of the *Escherichia coli* purine repressor." Biochemistry **41**(2): 511-20.
- Johnson, W. C., Jr. (1985). "Circular dichroism and its empirical application to biopolymers." Methods Biochem Anal **31**: 61-163.
- Jones, S., P. van Heyningen, et al. (1999). "Protein-DNA interactions: A structural analysis." J Mol Biol **287**(5): 877-96.
- Juo, Z. S., T. K. Chiu, et al. (1996). "How proteins recognize the TATA box." J Mol Biol **261**(2): 239-54.
- Kabsch, W. (1976). "LSQKAB program for Structural Analysis." Acta Crystallogr A **32**: 922-923.

- Kahmann, J. D., H. J. Sass, et al. (2003). "Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators." *Embo J* **22**(8): 1824-34.
- Kalodimos, C. G., N. Biris, et al. (2004). "Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes." *Science* **305**(5682): 386-9.
- Kieboom, J. and J. de Bont (2001). "Identification and molecular characterization of an efflux system involved in *Pseudomonas putida* S12 multidrug resistance." *Microbiology* **147**(Pt 1): 43-51.
- Kilstrup, M., L. M. Meng, et al. (1989). "Genetic evidence for a repressor of synthesis of cytosine deaminase and purine biosynthesis enzymes in *Escherichia coli*." *J Bacteriol* **171**(4): 2124-7.
- Kim, J. L., D. B. Nikolov, et al. (1993). "Co-crystal structure of TBP recognizing the minor groove of a TATA element." *Nature* **365**(6446): 520-7.
- Kim, Y., J. H. Geiger, et al. (1993). "Crystal structure of a yeast TBP/TATA-box complex." *Nature* **365**(6446): 512-20.
- Kissinger, C. R., D. K. Gehlhaar, et al. (1999). "Rapid automated molecular replacement by evolutionary search." *Acta Crystallogr D Biol Crystallogr* **55**(Pt 2): 484-91.
- Klein, J. R., B. Henrich, et al. (1991). "Molecular analysis and nucleotide sequence of the envCD operon of *Escherichia coli*." *Mol Gen Genet* **230**(1-2): 230-40.
- Koudelka, G. B. (1998). "Recognition of DNA structure by 434 repressor." *Nucleic Acids Res* **26**(2): 669-75.
- Koudelka, G. B. and P. Carlson (1992). "DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor." *Nature* **355**(6355): 89-91.
- Kumano, M., M. Fujita, et al. (2003). "Lincomycin resistance mutations in two regions immediately downstream of the -10 region of *lmr* promoter cause overexpression of a putative multidrug efflux pump in *Bacillus subtilis* mutants." *Antimicrob Agents Chemother* **47**(1): 432-5.
- Kuroda, M., T. Ohta, et al. (2001). "Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*." *Lancet* **357**(9264): 1225-40.
- Lakowicz, J. R. (1999). *Principles of Fluorescence Spectroscopy*. New York, Plenum Press.
- Laskowski, R. A., J. A. Rullmann, et al. (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." *J Biomol NMR* **8**(4): 477-86.
- Lavery, R. and A. Lebrun (1999). "Modelling DNA stretching for physics and biology." *Genetica* **106**(1-2): 75-84.
- Lebrun, A. and R. Lavery (1999). "Modeling DNA deformations induced by minor groove binding proteins." *Biopolymers* **49**(5): 341-53.
- Lee, E. H., C. Rouquette-Loughlin, et al. (2003). "FarR regulates the farAB-encoded efflux pump of *Neisseria gonorrhoeae* via an MtrR regulatory mechanism." *J Bacteriol* **185**(24): 7145-52.
- Lee, E. H. and W. M. Shafer (1999). "The farAB-encoded efflux pump mediates resistance of gonococci to long-chained antibacterial fatty acids." *Mol Microbiol* **33**(4): 839-45.
- Lewis, M., G. Chang, et al. (1996). "Crystal structure of the lactose operon repressor and its complexes with DNA and inducer." *Science* **271**(5253): 1247-54.

- Li, X. Z., L. Zhang, et al. (2004). "Efflux pump-mediated intrinsic drug resistance in *Mycobacterium smegmatis*." *Antimicrob Agents Chemother* **48**(7): 2415-23.
- Lilley, D. (1986). "DNA structure. Bent molecules--how and why?" *Nature* **320**(6062): 487-8.
- Lucas, C. E., Balthazar, J. T., Hagman, K. E., Shafer, W. M. (1997). "The MtrR repressor binds the DNA sequence between the *mtrR* and *mtrC* genes of *Neisseria gonorrhoeae*." *J. Bacteriol.* **179**(13): 4123-4128.
- Lundblad, J. R., M. Laurance, et al. (1996). "Fluorescence polarization analysis of protein-DNA and protein-protein interactions." *Mol Endocrinol* **10**(6): 607-12.
- Ma, D., M. Alberti, et al. (1996). "The local repressor AcrR plays a modulating role in the regulation of *acrAB* genes of *Escherichia coli* by global stress signals." *Mol Microbiol* **19**(1): 101-12.
- Ma, D., D. N. Cook, et al. (1995). "Genes *acrA* and *acrB* encode a stress-induced efflux system of *Escherichia coli*." *Mol Microbiol* **16**(1): 45-55.
- Martinez-Bueno, M., A. J. Molina-Henares, et al. (2004). "BacTregulators: a database of transcriptional regulators in bacteria and archaea." *Bioinformatics* **20**(16): 2787-91.
- Matthews, K. S. and J. C. Nichols (1998). "Lactose repressor protein: functional properties and structure." *Prog Nucleic Acid Res Mol Biol* **58**: 127-64.
- Mauro, S. A., D. Pawlowski, et al. (2003). "The role of the minor groove substituents in indirect readout of DNA sequence by 434 repressor." *J Biol Chem* **278**(15): 12955-60.
- McRee, D. E. (1999). "XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density." *J Struct Biol* **125**(2-3): 156-65.
- Meng, L. M. and P. Nygaard (1990). "Identification of hypoxanthine and guanine as the co-repressors for the purine regulon genes of *Escherichia coli*." *Mol Microbiol* **4**(12): 2187-92.
- Miura-Onai, S., M. Yabuta, et al. (1995). "Mutational study at Ser300 position of the *Escherichia coli* lactose repressor." *Biochem Biophys Res Commun* **209**(1): 126-30.
- Mowbray, S. L. and A. J. Bjorkman (1999). "Conformational changes of ribose-binding protein and two related repressors are tailored to fit the functional need." *J Mol Biol* **294**(2): 487-99.
- Nagadoi, A., S. Morikawa, et al. (1995). "Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain." *Structure* **3**(11): 1217-24.
- Nagadoi, A., K. Nakazawa, et al. (1995). "Conformational changes of purine repressor DNA-binding domain upon complexation with DNA." *Nucleic Acids Symp Ser*(34): 63-4.
- Namwat, W., C. K. Lee, et al. (2001). "Identification of the *varR* gene as a transcriptional regulator of virginiamycin S resistance in *Streptomyces virginiae*." *J Bacteriol* **183**(6): 2025-31.
- Napoli, A. A., C. L. Lawson, et al. (2006). "Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: recognition of pyrimidine-purine and purine-purine steps." *J Mol Biol* **357**(1): 173-83.

- Natsume, R., Y. Ohnishi, et al. (2004). "Crystal structure of a gamma-butyrolactone autoregulator receptor protein in *Streptomyces coelicolor* A3(2)." J Mol Biol **336**(2): 409-19.
- Nelson, H. C., J. T. Finch, et al. (1987). "The structure of an oligo(dA).oligo(dT) tract and its biological implications." Nature **330**(6145): 221-6.
- Orth, P., D. Schnappinger, et al. (2000). "Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system." Nat Struct Biol **7**(3): 215-9.
- Orth, P., D. Schnappinger, et al. (1999). "Crystal structure of the tet repressor in complex with a novel tetracycline, 9-(N,N-dimethylglycylamido)- 6-demethyl-6-deoxy-tetracycline." J Mol Biol **285**(2): 455-61.
- Otwinowski, Z., R. W. Schevitz, et al. (1988). "Crystal structure of trp repressor/operator complex at atomic resolution." Nature **335**(6188): 321-9.
- Pabo, C. O., A. K. Aggarwal, et al. (1990). "Conserved residues make similar contacts in two repressor-operator complexes." Science **247**(4947): 1210-3.
- Palumbo, J. D., C. I. Kado, et al. (1998). "An isoflavonoid-inducible efflux pump in *Agrobacterium tumefaciens* is involved in competitive colonization of roots." J Bacteriol **180**(12): 3107-13.
- Pan, W. and B. G. Spratt (1994). "Regulation of the permeability of the gonococcal cell envelope by the mtr system." Mol Microbiol **11**(4): 769-75.
- Parkinson, G., C. Wilson, et al. (1996). "Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface." J Mol Biol **260**(3): 395-408.
- Peng, W. T. and E. W. Nester (2001). "Characterization of a putative RND-type efflux system in *Agrobacterium tumefaciens*." Gene **270**(1-2): 245-52.
- Pflugrath, J. W. (1999). "The finer things in X-ray diffraction data collection." Acta Crystallogr D Biol Crystallogr **55**(Pt 10): 1718-25.
- Phillips, S. E., I. Manfield, et al. (1989). "Cooperative tandem binding of met repressor of *Escherichia coli*." Nature **341**(6244): 711-5.
- Ramos, J. L., M. Martinez-Bueno, et al. (2005). "The TetR family of transcriptional repressors." Microbiol Mol Biol Rev **69**(2): 326-56.
- Rebets, Y., B. Ostash, et al. (2003). "Production of landomycins in *Streptomyces globisporus* 1912 and *S. cyanogenus* S136 is regulated by genes encoding putative transcriptional activators." FEMS Microbiol Lett **222**(1): 149-53.
- Rojas, A., A. Segura, et al. (2003). "In vivo and in vitro evidence that TtgV is the specific regulator of the TtgGHI multidrug and solvent efflux pump of *Pseudomonas putida*." J Bacteriol **185**(16): 4755-63.
- Rolfes, R. J. and H. Zalkin (1988). "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**(36): 19653-61.
- Rolfes, R. J. and H. Zalkin (1988). "Regulation of *Escherichia coli* purF. Mutations that define the promoter, operator, and purine repressor gene." J Biol Chem **263**(36): 19649-52.
- Rouquette, C., J. B. Harmon, et al. (1999). "Induction of the mtrCDE-encoded efflux pump system of *Neisseria gonorrhoeae* requires MtrA, an AraC-like protein." Mol Microbiol **33**(3): 651-8.

- Sanchez, P., A. Alonso, et al. (2002). "Cloning and characterization of SmeT, a repressor of the *Stenotrophomonas maltophilia* multidrug efflux pump SmeDEF." *Antimicrob Agents Chemother* **46**(11): 3386-93.
- Sauer, R. T. (1995). "Minor groove DNA-recognition by alpha-helices." *Nat Struct Biol* **2**(1): 7-9.
- Sauer, R. T. (1996). "Lac repressor at last." *Structure* **4**(3): 219-22.
- Schumacher, M. A. and R. G. Brennan (2003). "Deciphering the molecular basis of multidrug recognition: crystal structures of the *Staphylococcus aureus* multidrug binding transcription regulator QacR." *Res Microbiol* **154**(2): 69-77.
- Schumacher, M. A., K. Y. Choi, et al. (1995). "Mechanism of corepressor-mediated specific DNA binding by the purine repressor." *Cell* **83**(1): 147-55.
- Schumacher, M. A., K. Y. Choi, et al. (1994). "Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices." *Science* **266**(5186): 763-70.
- Schumacher, M. A., K. Y. Choi, et al. (1994). "Crystallization and preliminary X-ray analysis of an *Escherichia coli* purine repressor-hypoxanthine-DNA complex." *J Mol Biol* **242**(3): 302-5.
- Schumacher, M. A., A. Glasfeld, et al. (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**(36): 22648-53.
- Schumacher, M. A., M. C. Miller, et al. (2001). "Structural mechanisms of QacR induction and multidrug recognition." *Science* **294**(5549): 2158-63.
- Schumacher, M. A., M. C. Miller, et al. (2002). "Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR." *Embo J* **21**(5): 1210-8.
- Sklenar, H., C. Etchebest, et al. (1989). "Describing protein structure: a general algorithm yielding complete helicoidal parameters and a unique overall axis." *Proteins* **6**(1): 46-60.
- Somers, W. S. and S. E. Phillips (1992). "Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by beta-strands." *Nature* **359**(6394): 387-93.
- Spronk, C. A., A. M. Bonvin, et al. (1999). "The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator." *Structure* **7**(12): 1483-92.
- Spronk, C. A., G. E. Folkers, et al. (1999). "Hinge-helix formation and DNA bending in various lac repressor-operator complexes." *Embo J* **18**(22): 6472-80.
- Stapleton, P., V. Adams, et al. (2004). "Characterisation of viridans group streptococci with different levels of Tet(M)-mediated tetracycline resistance." *Int J Antimicrob Agents* **24**(5): 439-43.
- Stout, G. H. and L. H. Jensen (1989). *X-Ray Structure Determination: A Practical Guide*, Joh Wiley & Sons, Inc.
- Swint-Kruse, L., C. Larson, et al. (2002). "Fine-tuning function: correlation of hinge domain interactions with functional distinctions between LacI and PurR." *Protein Sci* **11**(4): 778-94.

- Teran, W., A. Felipe, et al. (2003). "Antibiotic-dependent induction of *Pseudomonas putida* DOT-T1E TtgABC efflux pump is mediated by the drug binding repressor TtgR." *Antimicrob Agents Chemother* **47**(10): 3067-72.
- Thompson, J., Higgins D., Gibson, T. (1994). "Multiple Sequence Alignment." *Nature* **22**: 4673-4680.
- Trefzer, A., S. Pelzer, et al. (2002). "Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic." *Antimicrob Agents Chemother* **46**(5): 1174-82.
- Tronrud, D. E. (1997). "TNT refinement package." *Methods Enzymol* **277**: 306-19.
- Unger, B., G. Klock, et al. (1984). "Nucleotide sequence of the repressor gene of the RA1 tetracycline resistance determinant: structural and functional comparison with three related Tet repressor genes." *Nucleic Acids Res* **12**(20): 7693-703.
- van Holde, K. E., W. C. Johnson, Jr., et al. (1998). *Principles of Physical Biochemistry*. Up Saddle River, New Jersey, Prentice-Hall, inc.
- Veal, W. L. and W. M. Shafer (2003). "Identification of a cell envelope protein (MtrF) involved in hydrophobic antimicrobial resistance in *Neisseria gonorrhoeae*." *J Antimicrob Chemother* **51**(1): 27-37.
- Veal, W. L., A. Yellen, et al. (1998). "Loss-of-function mutations in the mtr efflux system of *Neisseria gonorrhoeae*." *Microbiology* **144** (Pt 3): 621-7.
- Weickert, M. J. and S. Adhya (1992). "A family of bacterial regulators homologous to Gal and Lac repressors." *J Biol Chem* **267**(22): 15869-74.
- Westbrock-Wadman, S., D. R. Sherman, et al. (1999). "Characterization of a *Pseudomonas aeruginosa* efflux pump contributing to aminoglycoside impermeability." *Antimicrob Agents Chemother* **43**(12): 2975-83.
- Wilson, H. R. and C. L. Turnbough, Jr. (1990). "Role of the purine repressor in the regulation of pyrimidine gene expression in *Escherichia coli* K-12." *J Bacteriol* **172**(6): 3208-13.
- Zalkin, H. a. N., P. (1996). *Biosynthesis of Purine Nucleotides*. Washington D.C., American Society of Microbiology Press.
- Zhang, R. G., A. Joachimiak, et al. (1987). "The crystal structure of trp aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity." *Nature* **327**(6123): 591-7.
- Zheng, L., U. Baumann, et al. (2004). "An efficient one-step site-directed and site-saturation mutagenesis protocol." *Nucleic Acids Res* **32**(14): e115.