CRYSTALLOGRAPHIC STUDIES ON THE MULTIDRUG TRANSPORTER ACTIVATION PROTEIN N-TERMINUS FROM *BACILLUS SUBTILIS*

by

Michael Hiday Godsey

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

December 2001

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. thesis of

Michael Hiday Godsey

has been approved

[redacted]

Professor in charge of thesis

[redacted]

Member

[redacted]

Member

[redacted]

Member

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

TABLE OF ABBREVIATIONS

MDR	Multidrug resistance
TM	Transmembrane
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
ABC	ATP-binding cassette
MFS	Major facilitator superfamily
RND	Resistance/nodulation/division superfamily
DMT	Drug/metabolite transporter superfamily
MATE	Multidrug and toxic compound extrusion superfamily
MET	Multidrug endosomal transporter
NBD	Nucleotide-binding domain
SMC	Structural maintenance of chromosomes
mRNA	Messenger ribonucleicacid
PMF	Proton motive force
RNAP	Ribonucleicacid polymerase
HTH	Helix-turn-helix
G	Guanine
C	Cytosine
A	Adenine
T	Thymine
PCR	Polymerase chain reaction
LB	Luria-Bertani
TPP^+	Tetraphenylphosphonium
MIR	Multiple isomorphous replacement
MR	Molecular replacement
MAD	Multiwavelength anomalous diffraction
XAS	X-ray absorption spectrometry
DM	Density modification
FA	Fluorescence anisotropy
FP	Fluorescence polarization

ACKNOWLEDGMENTS

I wish to thank E. E. Zheleznova-Heldwein for sharing the coordinates of BmrR prior to publication. I must acknowledge Maria Schumacher for instruction in the finer and grosser points of all aspects of X-ray crystallography. I also want to thank the other members of the lab, Greg Allen, Joy Huffman, Kate Hoffmann, Dave Murray and Minsun Hong, for creating a challenging and interesting work environment. Pete von Hippel, Rick Dahlquist and Brian Matthews of the University of Oregon were positive influences during my time there, and all were helpful and encouraging in my transfer to OHSU. Mostly, I wish to thank Dick Brennan for being friendly from our first meeting, accepting me into his laboratory, and giving me a chance to learn.

Abstract

Multidrug resistance (MDR) is a growing human health challenge. One mechanism of MDR in bacteria is the overproduction of multidrug transporters. Multidrug (MD) transporters comprise a superfamily of pumps, which are able to expel a wide variety of toxic compounds from cells. They have been found in organisms from bacteria to humans, and it is believed that they probably exist in some form in all living organisms except for viruses.

Overproduction of MD transporters is deleterious to cells under normal growth conditions, because their substrates include some essential metabolites. However, overproduction of these pumps by leads to the MDR phenotype. Cells in which these pumps are overexpressed or induced by substrates are able to survive normally toxic levels of drugs. Thus, regulation of MD transporters is likely to be a key component of cell survival.

Transcriptional regulators bind drugs to control the production of MD transporters in response to chemical challenges. Members of the MerR family of regulators control transcription of several MD transporters in *Bacillus subtilis*. Other MerR proteins respond to signals such as heavy metals, large antibiotics, soybean nodulation factors, and oxidative stress. This variation in signal responsiveness is made possible through large differences in the C-terminal effector-binding regions, while the N-terminal DNA-binding domains remain homologous across the family.

In *Bacillus subtilis*, BmrR and BltR control the transcription of two multidrug transporter genes, *bmr* and *blt*, respectively. It was believed that they are further controlled by a global regulator, Mta. All three regulators are MerR family members. Mta was identified through a stop mutation, which produces a protein that contains only the first 109 amino acyl residues of the Mta protein. This truncated mutant was named MtaN, for Mta N-terminus. MtaN constitutively activates transcription of *bmr* and *blt*,

its own transcription, and that of a fourth gene, *ydfK*, which encodes a putative membrane protein.

The structure of a BmrR-DNA-drug complex reveals a fully bound and activated form of a MerR family member. However, structural changes occurring in the protein to allow DNA-binding and activation were unknown, because no unbound or inactivated structure of a MerR family member was yet available. This thesis reports the first structure of an apo form of a MerR family member, that of MtaN.

A comparison of the structures of MtaN and BmrR-DNA-drug reveals a significant difference in the dimeric conformations of the proteins. The difference is suggestive of a mechanism of DNA-activation whereby a MerR protein exists in an untwisted form until it binds DNA and activator, at which time changes are transmitted from one drug binding site to the opposite monomer through conformational changes in the long dimerization helix. These changes lead to a twisting and contraction of the dimer, causing an undertwisting and shortening of the DNA, bringing the –10 and –35 elements of the promoter into appropriate orientation for open promoter complex formation.

Chapter 1

INTRODUCTION

1.1 Multidrug Resistance

As the use of antibiotics and other antimicrobials has proceeded through the second half of the last century, a large number of treatment-resistant strains of human pathogens has emerged. Examples of this include some of the deadliest microbes ever encountered: *Mycobacterium tuberculosis*, *Plasmodium falciparum* (malaria), human immunodeficiency virus (HIV), and *Staphylococcus aureus*, to name a few. Most of these resistant strains became immune to the first-line drugs used to treat them through the process of evolution forced by the selective pressure of the chemotherapeutic agents. However, in most cases, there exist second- or third-line treatments for these diseases, which remain effective.

It is theorized that because all antibiotics in common use are derived from natural sources (mostly fungi) that resistance mechanisms to these compounds have evolved over the ages. Humans may have simply increased the selective pressure and allowed these mechanisms to become more widespread, rather than the microbes inventing new resistance mechanisms on an evolutionarily short time scale. To explore and exploit this possibility, some new, synthetic antibiotics are in preparation, which may help to turn the tide again in favor of medical science (Muller *et al.* 1999). Initial results are very promising. Unfortunately, it is probable that bacteria will also eventually develop resistance to these synthetic drugs as well. As time goes on, strains resistant to single

treatments will acquire resistance to other treatments, and become increasingly difficult and expensive, if not impossible, to treat. Already in hospitals, strains of *S. aureus* have appeared which are resistant to all antibiotics currently in use (Lyon *et al.* 1987; Tenover 1999). The emergence of multiple-antibiotic resistant strains of other potent human pathogens has also been documented in hospitals around the world (reviewed in (Travis 1994)).

Multidrug resistance (MDR) can be loosely defined as the ability of cells to withstand normally fatal levels of a range of toxic compounds. In discussions of MDR, these normally fatal doses are the standard concentrations for effectiveness, and the toxic compounds are referred to as drugs. The role of MDR in human medicine is becoming increasingly significant. Thus, understanding the molecular mechanism of MDR is also of increasing importance

It is well known that microbes often exchange genetic material, including those genes responsible for drug resistance. It is also appreciated that mutations occur frequently in pathogens, some of which may confer resistance to a drug. Either of these events may be able to produce resistance in an organism. However, when MDR strains appeared, it was seen as highly unlikely that many mutations, or several gene-transfer events, would happen simultaneously that would confer resistance to a variety of treatments. Another possibility is that one mechanism is responsible for the failure of many treatments.

Biochemical mechanisms of drug-resistance

There are many biochemical mechanisms of drug resistance, such as (Figure 1.1): (1) drug inactivation before it reaches its cellular target; (2) reduced metabolic activation of drug; (3) alteration of the cellular target of the drug; (4) duplication of target function by a non-target protein; (5) increase in the number of target sites; (6) increased repair of drug damage; (7) sequestration of drug; (8) decreased uptake of drug; and (9) increased efflux of drug. These mechanisms fall into a few general schemes.

The first general scheme is the reduction of the activity of drug in the cell by either inactivating it or never activating it. Drug inactivation is exemplified by the β-lactamases, enzymes that cleave an important bond in the ring structures of β-lactam antibiotics, which include the penicillins and cephalosporins. This mechanism of resistance was found in cells that were frozen even before the medicinal development of penicillin (Davies 1994; Travis 1994). Some drugs require metabolic processes to become toxic, such as some purine anti-metabolites. Prontosil, an early sulfa drug, is non-toxic to bacteria until it is metabolized into sulfanilamide by an animal. Either metabolic inactivation or reduction of the metabolic processes that activate drugs will lead to a lower level of active drug.

A second general scheme of resistance is circumventing the action of the drug by using an active site insensitive to drug. This is done either by alteration (mutation or modification) of the target site or by the use of an alternate pathway. When a drug has a very specific target, a minor change in the active site of the target may lead to resistance without significantly sacrificing activity of the protein. In *E. coli* DNA gyrase, changes

of this type were found to confer resistance to nalidixic acid and novobiocin (Reynolds 1984), and in RNA polymerase subtle changes were found to confer resistance to rifampicin (Rabussay *et al.* 1969). Sometimes, resistance is gained through the increased use of alternative proteins or pathways to achieve the same metabolic action as the drug target would normally accomplish, as is seen in the purine salvage pathways of protozoan parasites such as *Toxoplasma gondii* and *Leishmania donovani* (Vasudevan *et al.* 2001) in which inactivation of one salvage enzyme or transporter can be compensated for by increased activity of another. In bacteria, trimethoprim inhibits cellular dihydrofolate reductase (DHFR) activity, however a plasmid-encoded DHFR enzyme was found to differ significantly from the trimethoprin target and confer resistance (Smith *et al.* 1984; Amyes *et al.* 1989).

The third scheme involves simply overpowering the drug effects. This can be done by increasing production of the drug target. For example, some methotrexate-resistant cell lines are found to overproduce DHFR, the methotrexate target (Nunberg *et al.* 1978; Alt *et al.* 1992). This can be achieved through several genetic mechanisms ranging from gene amplification to derepression. Another mechanism to overpower the drug is to increase the repair activities that counter the drug effects. Many drugs target the DNA bases by modifying them, however these effects can be sometimes be overcome by an increase in cellular DNA-repair systems (Kessel 1986).

The fourth scheme is to lower the level of drug inside the cell by changes in the intake and efflux. Changes in expression or action of a drug import protein or changes in

the cell wall or membrane to make them less permeable to drug would lower the level of drug being introduced to the interior of the cell. (Mutations in the exterior proteins of viruses are very common, including rhinoviruses (the common cold) and HIV. While these viral proteins do not influx drugs, they are often targets of vaccines, and their frequent mutations make the development of effective vaccines more difficult.) The alternative to reduced intake is increased efflux, which is a common phenomenon in MDR (Nikaido 1996; Paulsen *et al.* 1996; Putman *et al.* 2000). Overexpression or increased activation of drug-efflux pumps can also lead to lower levels of drugs in the cell. MDR transporters in *Bacillus subtilis* will be discussed in this thesis. However, the first example of MDR described also used this type of mechanism.

Multidrug transporters

The study of MDR began with the discovery of human tumor cells that had become resistant to a spectrum of anti-cancer compounds. Ironically, many chemotherapeutic agents, especially in cancer treatment, are also mutagenic and can facilitate mutations which give rise to resistance (Stark 1986a; Stark 1986b). It was found that these cells expressed a membrane protein that led to a lower concentration of chemotherapeutic agents in the cells (Juliano *et al.* 1976). P-glycoprotein was identified as the transporter that was able to confer resistance to multiple anti-cancer drugs when overexpressed (Kartner *et al.* 1983a). P-glycoprotein inhibitors were found to restore susceptibility. This was the first example of a single protein conferring the MDR phenotype, and this was confirmed in tumor cell lines from hamsters and mice (Kartner *et al.* 1983a; Kartner *et al.* 1983b; Ling *et al.* 1983).

Since the discovery of role of P-glycoprotein in MDR, it has become clear that these transporters are a common mechanism of MDR. There are now five superfamilies of MDR transporters, which were reviewed by Saier and Paulsen (Saier *et al.* 2001). They are: the ATP-binding cassette (ABC) superfamily, which includes P-glycoprotein and other proteins with 12 transmembrane (TM) domains in a dimeric or pseudodimeric structure; the 12 or 14 TM major facilitator superfamily (MFS), including Bmr and Blt from *B. subtilis*; the 8 TM resistance/nodulation/division (RND) superfamily, ubiquitous in bacteria, archaea and eukaryotes; the 10 TM drug/metabolite transporter (DMT) superfamily; the 12 TM multidrug and toxic compound extrusion (MATE) superfamily; and the 4 TM multidrug endosomal transporter (MET) superfamily. In several of these families, the number of TM domains is only putative, and further studies are needed to confirm the predicted topology. An in-depth examination of all these superfamilies is beyond the scope of this work, but they do share the ability to transport chemicals that are only loosely related (Nikaido 1996; Paulsen *et al.* 1996; Putman *et al.* 2000).

Broad specificity of multidrug transporters

In addition to the medicinal importance of multidrug transporters, they are also scientifically interesting for their ability to transport a variety of substrates. Most proteins interact strongly only with one or a few specific substrate(s), and weakly with others. Some proteins act on a specific moiety in a substrate and tolerate substitution of groups not directly involved in binding or catalysis. MDR transporters, on the other hand, are able to recognize, bind, and transport a relatively wide spectrum of organic compounds, including antibiotics, dyes, drugs, and other hydrophobic cations.

Many transporters that are closely related to MDR transporters have specific substrates, such as most of the ABC transporters (Fath *et al.* 1993). The histidine permease from *Salmonella typhimurium*, and the maltose transporter of *Thermococcus listoralis* are discussed below. Members of this family also include protein and peptide transporters and several are known to transport specific antibiotics (Putman *et al.* 2000). This suggests that MDR transporters might have evolved loose specificity from more specific ancestors, perhaps as part of the microbial arms race. Faced with diverse chemical attacks from numerous competing organisms, MDR is a potent defense mechanism, allowing resistance to a variety of toxic compounds through a single mechanism. Cells with this capability would be able to survive in environments unsuitable to their sensitive species-mates. However, it is also possible that MDR transporters evolved for other reasons. Rather than evolving as defensive systems against exogenous toxic chemicals, they may have evolved to expel endogenous toxins. While classified as a MDR transporter (Ahmed *et al.* 1995), Blt from *B. subtilis* is also known to transport spermidine (Woolridge *et al.* 1997), and it is co-regulated with BltD, a spermine/spermidine acetyltransferase (Woolridge *et al.* 1999), suggesting that its primary function is in polyamine transport.

Mechanism of multidrug transporters

A simple model of transporter action is the membrane-bound transporter binding a substrate in the cytosol, undergoing a conformational change to expose the substrate to the exterior of the cell, releasing the substrate to the exterior, and changing conformation back to a form able to bind substrate in the interior of the cell again. However, the nature

of the substrates of MDR transporters implies that perhaps they are not classical pumps, as described above. MDR transporter substrates, being mainly hydrophobic cations, partition preferentially to the leaflets of the membrane, an area near where the charged heads and hydrophobic tails of the membrane lipids connect (Raviv *et al.* 1990; Higgins *et al.* 1992; Ueda *et al.* 1997). In a mechanism known as the "flippase" model, MDR transporters are found to bind their substrates on the inner leaflet of the membrane and "flip" them to the outer leaflet, which results in the removal of substrate from the cytosol (Higgins *et al.* 1992). Evidence of this has been mounting in recent years (Bolhuis *et al.* 1996; Bolhuis *et al.* 1997; van Veen 2001). The current model of the ABC pumps, based on LmrA, is a two-cylinder engine (Figure 1.2), with two drug binding sites per functional complex (van Veen *et al.* 2000). This is in agreement with the known dimeric or pseudo-dimeric nature of the ABC superfamily. The predominant features of this model are a high-affinity drug binding site on the inner leaflet and a low-affinity binding site on the outer leaflet. The net effect of the ATP hydrolysis is to move the drug from the high-affinity site to the low-affinity site.

The structure of an ABC transporter, *E. coli* MsbA, has recently given limited support to the flippase model (Chang *et al.* 2001). MsbA is a transporter that moves lipid A, which is necessary for the bacterial outer membrane (Zhou *et al.* 1998). The structure of this membrane protein was solved at 4.5 Å resolution, which is sufficient to show a large entrance from the inner leaflet of the membrane to a central cavity located between the two monomers (Figure 1.3). This cavity is probably large enough to accommodate most drug molecules (although MsbA is not an MDR transporter). The structure also

suggests a general mechanism of hydrophobic molecule transport by ABC transporters. A large contact area between the intracellular ATP-binding cassettes (ABC) and the membrane spanning regions could couple ATP hydrolysis to global structural changes and substrate binding. Binding of substrate in the inner chamber of the membrane-spanning region could trigger the intracellular ABC to hydrolyze bound ATP. Hydrolysis might allow the nucleotide binding domains to interact and move closer together. This movement would close the opening of the chamber to the inner leaflet and allow the formation of an opening to the outer leaflet of the membrane or an external low-affinity binding site. After the substrate has flipped to the outer leaflet and diffused away or been expelled into the extracellular space, the chamber would close and signal the release of ADP by the ABC. The crystal structure is of insufficient resolution to identify or characterize substrate-binding sites, and it does not give evidence of four possible binding sites per dimer.

More is known structurally about the nucleotide binding domains (NBD) of the ABC transporters (Schneider *et al.* 1998). ATP molecules bind and are hydrolyzed by these domains. The crystal structures of three NBD from two ABC transporters and a DNA double-strand break repair protein are available (Figure 1.4). The crystal structures of HisP (Hung *et al.* 1998), a component of the histidine permease from *Salmonella typhimurium*, and MalK (Diederichs *et al.* 2000) from the maltose transporter of *Thermococcus listoralis*, were solved to 1.5 Å and 1.9 Å resolution, respectively. The Rad50 ATPase domain from *Pyrococcus furiosus* (Hopfner *et al.* 2000), involved in DNA double-strand break repair, was found to have a similar structure suggesting that this

catalytic domain is functional in several different contexts. The structure of SMC (structural maintenance of chromosomes) (Lowe *et al.* 2001) from *Thermatoga maritima*, and the partial structure of MukB (van den Ent *et al.* 1999), a SMC from *E. coli*, also show examples of this fold.

While the nucleotide binding domains of these proteins do not directly bind any drugs for transport, their structures shed light on the mechanism of efflux used by these ATP-dependent transporters. In particular, the Rad50 ATPase domain, which was solved in the presence and absence of ATP, reveals the role of the conserved ABC signature motif in ATP binding and dimerization. The ATP-induced dimer buries both bound ATP molecules and shows that the γ-phosphate of ATP bound by one monomer is contacted by the signature motif of the opposite monomer. This contact may serve as a sensor of ATP in the opposite monomer and tie hydrolysis of ATP to structural changes, and ultimately ligand transport, by the ABC transporters (Hopfner *et al.* 2000). This would be consistent with the 2-cylinder model described for LmrA (van Veen *et al.* 2000).

Less is known about the workings of the other families of MDR transporters. The other widely studied superfamily of transporters is the major facilitator superfamily (MFS), which includes Bmr and Blt from *B. subtilis* and NorA and QacA of *S. aureus*. The >1000 sequenced members of this family have either 12 or 14 transmembrane helices and are found in bacteria, archaea, and eukaryotes (Pao *et al.* 1998; Saier *et al.* 1999). Bmr, NorA and Blt are members of the 12-TM subset of this family. All three are inhibited by the known P-glycoprotein inhibitor, reserpine, and can transport a similar

spectrum of compounds, including rhodamine, ethidium and fluoroquinolones. All known MFS transporters involved in drug efflux are H^+ antiporters. The action of these proteins is dependent upon and driven by the proton motive force (PMF), the proton gradient across the membrane.

1.2 Regulation of transcription

Prokaryotes regulate transcription efficiently in several ways. One method is clustering genes with related functions into groups called operons. Because many metabolic processes require multiple steps and multiple enzymes, this grouping means that one regulation mechanism can be used for the whole pathway, instead of requiring separate regulation for each of the related genes. The *blt* and *bmr* gene groups in *B. subtilis* are examples of this type of genomic organization, as is the *lac* operon in *E. coli*. The *blt* operon contains only two known genes, *blt and bltD*, both of which are controlled by BltR, the product of a nearby gene. The *bmr* gene is upstream of the *bmrR* gene and downstream of the *bmrU* gene, all of which are under the regulation of BmrR. The *lac* operon is described in more detail below.

Not surprisingly, transcriptional regulation in bacteria is achieved through protein-DNA interactions and requires cis-acting elements in addition to trans-acting proteins. The cis-acting elements are parts of the DNA called promoters and operators. These DNA sequences may bind regulatory proteins (these are called operators) or increase the likelihood of RNA polymerase (RNAP) binding (these are called promoters). Without promoters, the RNAP would not specifically recognize transcription start sites. The trans-acting regulatory proteins often respond to environmental signals.

The *lac* repressor is a transcriptional regulator of genes involved in the metabolism of lactose in *E. coli*. It is a repressor that specifically binds DNA near the genes for lactose metabolism: import protein (*lacY*), β-galactosidase (*lacZ*), which breaks lactose into glucose and galactose, and β-galactoside transacetylase (*lacA*), which moves an acetyl group from acetyl-coA to β-galactosides (Jacob *et al.* 1961). In the absence of lactose, the repressor binds the operator (*lacO*) and blocks RNAP, thus RNAP is unable to transcribe the genes for the proteins of lactose metabolism. However, when lactose is present it binds to the repressor and reduces the affinity of the repressor for the operator site. In doing so, the repressor dissociates from the operator, and RNAP can effectively transcribe the genes for lactose metabolism, thus preparing the cell to use lactose as a carbon and energy source. MarR (*E. coli*) and QacR (*S. aureus*) are other examples of repressors which dissociate from the DNA when signal molecules are bound, which are discussed elsewhere in this work.

An alternative to the repressor model is that of the activator. Activators can stimulate specific genes, like BmrR acts on *bmr*, or activate globally, such as MarA (Martin *et al.* 1996). Activators bind to sequences near genes and promote transcription of those genes. MarA activates a number of genes involved in multidrug resistance and stress response in *E. coli* without binding to any signal molecules; instead, levels of MarA are controlled by the multidrug-sensing repressor MarR (Martin *et al.* 1996). In contrast, BmrR always binds specifically to the *bmr* promoter and activates transcription only when a ligand (co-activator or drug) is bound (Ahmed *et al.* 1994). While both are

examples of activators, they highlight the different approaches that are used to regulate genes.

Regulation of MDR genes at the transcriptional level

In cancer cells, multidrug resistance is often due to an overexpression of the *mdr* gene, which produces P-glycoprotein (Higgins *et al.* 1985; Higgins *et al.* 1986; Cole *et al.* 1992). Through increased expression and subsequent gene amplification events, the tumor cells acquire resistance to various anti-cancer compounds (Roninson *et al.* 1986; Shen *et al.* 1986). In *B. subtilis*, overexpression of Bmr and Blt also leads to increased resistance to toxic compounds (Ahmed *et al.* 1994; Putman *et al.* 2000). High levels of expression of MDR genes would therefore seem to be advantageous to the cell. However, because MDR pumps have a wide range of specificity, they probably transport some essential metabolites in addition to toxins, thus making high levels of constitutive expression deleterious to the cell under normal growth conditions. In addition, high levels of expression may be damaging to the integrity of the membrane. Therefore, these genes must be under some form of regulation. Indeed, it has been found that many bacterial MDR pumps are under the direct control of specific transcriptional regulators (Grkovic *et al.* 2001b). Furthermore, it has also been found that many of these regulators bind some of the same compounds that their respective pumps expel (Ahmed *et al.* 1994; Grkovic *et al.* 1998; Alekshun *et al.* 2001). An exception to this is the small multidrug resistance (SMR) family of transporters, for which no evidence of transcriptional regulation has yet been found (Grkovic *et al.* 2001b).

marA upon drug binding, thereby indirectly activating genes under the control of MarA. AcrR is a TetR family repressor, which does not appear to respond to drugs, but rather modulates the levels of *acr* transcription (Ma *et al.* 1996). SoxS and Rob are also activators of the *acr* and *tolC* genes, however they are more likely to be general stress-response than multidrug-response proteins (Ariza *et al.* 1995). Detached from those systems, EmrR is a repressor of the *emr* locus and has homology to MarR. EmrR was also found to bind substrates of the EmrAB pump, which it regulates, leading to derepression (Lomovskaya *et al.* 1995; Brooun *et al.* 1999) of *emrAB* and the MDR phenotype. Overall, *E. coli* has perhaps the best-understood regulatory system of MDR genes in bacteria. The structures of several of these proteins have recently become available and are reviewed in appendix A.

In *Bacillus subtilis*, a gram-positive bacteria, MDR is under the control of several MerR family proteins (Figure 1.6). BmrR and BltR control transcription of the *bmr* and *blt* genes, respectively (Ahmed *et al.* 1994; Ahmed *et al.* 1995). Both genes are also under the apparent influence of a global regulator, Mta, another MerR family member (Baranova *et al.* 1999). Bmr and Blt are similar in that they are both members of the 12-TM family of MFS proteins, share 51% sequence identity, and have a similar set of substrates. However, they are divergently regulated under normal metabolic circumstances. Bmr is expressed under normal growth conditions, while Blt is not (Ahmed *et al.* 1995). Furthermore, *bmr* is inducible by the drug substrates of Bmr through the action of BmrR (Ahmed *et al.* 1994), while no inducers of BltR have been found. In the *blt* operon, in addition to the Blt transporter, is a gene (*bltD*) that encodes an

acetyltransferase involved in spermidine degradation (Woolridge *et al.* 1999). Blt is also known to transport spermidine, suggesting that this may be the natural role for this transporter, and that multidrug transport is a secondary function (Woolridge *et al.* 1997). The discovery of Mta is discussed in a later section specific to MerR proteins, but the truncated form of Mta (MtaN) is able to activate transcription of both *bmr* and *blt* constitutively (Baranova *et al.* 1999). The inducers of Mta are unknown.

DNA-binding proteins

In order for activators and repressors to function, they must be able to recognize and bind specific DNA sequences. Protein-DNA recognition and binding follows principles similar to those of protein-protein interactions. The principles of molecular recognition are based on complementarity of shape, charge, and hydrogen bond donor and acceptor patterns within interfaces. As in protein folding and protein-protein interactions, the major sources of energy for binding are the hydrophobic effect and the satisfaction of hydrogen bond donors and acceptors.

Molecular recognition by proteins is represented by one of two models: a lock and key or an induced fit (Janin 1995; Jones *et al.* 1996). In the lock and key model, a protein has a precise shape, charge and bonding pattern that allow it to bind to only complementary DNA, protein, or other ligand. This is exemplified by the immunoglobulins (Davies *et al.* 1996), proteins designed to be perfectly complementary to antigens, although even in this example, small motions and water molecules are sometimes observed in the binding sites. This model explains specificity very well because the protein and ligand fit together precisely. However, this model does not

explain the ability of proteins to recognize diverse ligands (a common theme in multidrug-binding proteins) or to accommodate binding to non-optimal molecules.

Proteins can experience changes upon ligand binding allowing them to stabilize or form binding elements. These changes are examples of induced fit. DNA-binding proteins are often found to have different conformations on and off DNA, such as λ-Cro (Anderson *et al.* 1981; Brennan *et al.* 1990), which was found to have large rotations upon DNA binding. *E. coli* PurR, the purine repressor, also changes upon DNA and ligand binding; the binding of ligand, a purine base, alters the shape of the protein, and changes its ability to bind DNA by allowing the formation of hinge helices that bind to the minor groove towards the center of the operator sequence (Schumacher *et al.* 1994; Schumacher *et al.* 1995). In this case, the ligand acts as a co-repressor, and binding of the ligand allows the protein to bind DNA and therefore repress transcription of the genes for purine biosynthesis.

B-DNA has a regular structure with several characteristic properties. First, the overall charge of DNA is strongly negative due to the charged nature of the phosphate backbone. To exploit this, many DNA-binding proteins have positively charged patches to balance the overall charge of the complex. The phosphate backbone also provides a large number of potential hydrogen bonding partners, but electrostatic and hydrogen-bonding contacts to the backbone are non-specific because they could presumably form on any stretch of B-DNA.

Watson-Crick hydrogen bonding between bases on opposite strands of B-DNA results in a wide major groove (~14 Å) and a narrow minor groove (~7 Å), both of which allow access to potential hydrogen-bonding groups of the bases. These major and minor groove hydrogen-bonding patterns are specific to the bases found in the DNA, thus they are the primary determinants of recognition by DNA-binding proteins (Seeman *et al.* 1976). The four base pairs have different patterns (Figure 1.7). A guanine base in a GC base pair presents two hydrogen bond acceptors in the major groove, and one acceptor and one donor in the minor groove. The cytosine has a hydrogen-bond donor and a hydrogen atom in the major groove, and an acceptor the minor groove. In an AT base pair, thymine shows a methyl group and a hydrogen-bond acceptor in the major groove, and another acceptor in the minor groove. The adenine has an acceptor and a donor in the major groove and an acceptor and a hydrogen atom in the minor groove. When the AT base pair is switched to TA, the pattern of donors and acceptors reverses direction, and the same happens when GC switches to CG. This provides a recognizable pattern in the major groove to distinguish between the four base pairs. However, in the minor groove, the pattern for CG and GC base pairs are the same, as are the AT and TA patterns.

In order to recognize the hydrogen-bond donors and acceptors in the grooves of DNA, proteins must present a complementary set of donors and acceptors. This complementarity does not have to be exact, only good enough to provide the necessary specificity. There are multiple hydrogen-bond donors and acceptors throughout a protein chain. Backbone nitrogen atoms are donors and the backbone carbonyl oxygens are acceptors. However, these are usually involved in secondary structure elements and, like

the backbone of DNA, are not specific, *per se* (although the variation of protein shapes does add an element of specificity). More significantly, proteins also have hydrogen bond donors and acceptors on their side chains. The side chains of asparagine, glutamine, serine and threonine can serve as hydrogen bond donors or acceptors. Those of tryptophan and arginine may donate hydrogen-bonds at any pH. The hydrogen-bonding abilities of lysine, tyrosine, aspartic and glutamic acids and histidine vary with pH, due to their ionizable functional groups. At neutral pH, tyrosine and lysine are usually protonated and can serve as donors, and the acids are usually deprotonated and can serve as acceptors. The pK_a of the histidine side chain is 6.5, and its local environment plays a large part in its protonation state and therefore its hydrogen-bonding potential. Water molecules are also often found mediating hydrogen-bonds in the protein-DNA interface (Janin 1999).

Through the complementarity principles discussed above, a set of amino acyl residues can match a set of DNA bases to form a non-covalent complex. The major groove of B-DNA is wide enough to accept an α-helix or two β-strands, and proteins have developed structural motifs designed to place these elements into the major groove. Common examples of these motifs include: the zinc finger motif, which bind a zinc ion to stabilize a conformation capable of binding DNA (Laity *et al.* 2001); leucine zippers, which resemble two chopsticks wrapping around each other while trying to pick up the DNA (Busch *et al.* 1990); the helix-loop-helix motif, which may have a long zipper region as well, but has a variable loop N-terminal to the helices that insert into the DNA (Phillips 1994); and the helix-turn-helix (HTH) motif (Brennan *et al.* 1989), which will be discussed in more detail below. There are also proteins which use β-sheet structures to

bind the major groove, and proteins which bind primarily or entirely to the minor groove (Bewley *et al.* 1998). The topic of this work is a winged-helix protein (Brennan 1993; Gajiwala *et al.* 2000a), which contains the HTH motif and a β-sheet structure.

The HTH motif

The most common DNA-binding motif in bacteria is the helix-turn-helix. This motif was first observed in the crystal structures of λ-Cro (Anderson *et al.* 1981), CAP (McKay *et al.* 1981), and λ-repressor (Pabo *et al.* 1982), and the pattern was recognized very soon after the solution of these structures (Sauer *et al.* 1982). Sequence analyses of these and other proteins has led to the fairly reliable prediction of HTH motifs in proteins (Dodd *et al.* 1987; Dodd *et al.* 1990). The HTH consists mainly of two α-helices at an approximately 120 degree angle to each other joined by a 4-residue turn. The helices are packed against each other to bury a hydrophobic region between them. Steric constraints limit the possibilities of residues in certain positions. Using a numbering system based on 22 residues in the HTH motif, the first residue of the turn (residue 9) is usually a glycine because of the ability of that amino acyl residue to adopt conformations inaccessible to residues with side chains, residues 4 and 15 should be uncharged because they are buried, residues in the helices (3-8 and 15-20) should not be proline, and residue 5 should not be β-branched because of steric clashes (Brennan *et al.* 1989). Of course, there are exceptions to these principles, but these and other generalities have been used fairly reliably in the identification of new HTH proteins. As an example, MtaN, the subject of this thesis, follows all of these rules.

In prokaryotic HTH proteins, binding usually occurs by homodimerization, which results in two helices protruding from the same face of the dimer. This arrangement was seen in the first HTH proteins, which were solved without bound DNA. Because they were close to parallel and approximately 34 Å apart from each other, it was proposed that these helices would insert into consecutive major grooves and provide DNA recognition elements (Anderson *et al.* 1981). For this reason, the second helix of the HTH is often referred to as the "recognition" helix (Figure 1.8). This prediction was borne out by the structure solution of a HTH protein in complex with DNA (Anderson *et al.* 1987). The 434 repressor was solved to 3.2 Å resolution bound to DNA and showed that, as predicted, the recognition helices inserted themselves into consecutive major grooves of the DNA.

Since the discovery and characterization of the HTH motif, a number of variations of, and additions to, the basic framework of this structural element have been found. Variations of HTH motifs are found in eukaryotic homeodomains and POU-specific domains. Another variation is the winged-helix motif (Brennan 1993; Gajiwala *et al.* 2000a) (Figure 1.9). First recognized in the structure of the hepatocyte nuclear factor 3γ (HNF-3γ) in complex with DNA, this motif adds β-sheet structure and DNA-contacting loops to the HTH motif (Clark *et al.* 1993). The overall structure of the classic winged-helix motif is α/β with the topology H1-S1-H2-H3-S2-W1-S3-W2, where H is helix, S is β-strand and W is wing. Helix 3 is the recognition helix of the HTH, flanked by wing 1 and wing 2. In many winged-helix proteins, the HTH differs from the canonical HTH by possessing a longer turn, which allows more flexibility between the helices. Insertions in

the turn are found in POU-specific domains and homeodomains, although the helices of the HTH motifs retain the HTH relative orientations. The additions to and insertions in the HTH of a winged-helix protein are necessary because proteins, to ensure proper binding, must use extensive contacts to the DNA, which cannot be achieved through a single helix (Brennan 1993). Other contacts must be involved, either by dimerization or other elements of the domain.

While the unifying theme of the winged-helix domain is simple, the deviations from the theme are numerous. There are examples of winged-helix proteins that bind DNA as monomers, homodimers and heterodimers (Gajiwala *et al.* 2000a). The secondary structure topology described above is also variable. Most have longer turns than the classic four-residue turn found in the regular HTH, and the angles between the helices of the HTH can vary from 100° to 150°, while in regular HTH domains they differ little from the average 120°. The loops that are the wings can vary in size, location, DNA and protein contacts, and even secondary structural elements. In one winged-helix protein, RFX1, the "recognition helix" sits near the DNA backbone, while the wing binds in the major groove (Gajiwala *et al.* 2000b). While there are numerous variations on the structure and function of winged-helix domains (Gajiwala *et al.* 2000a; Wolberger *et al.* 2000), the general theme is the use of loops flanking the HTH to contact the DNA. The MerR family has yet another variation on the winged-helix domain, discussed later. Based upon sequence homology, it is likely that all MerR family member proteins will display a similar form of the winged-helix.

1.3 MerR proteins

Background, discovery, function

MtaN and BmrR are members of the MerR family of proteins, named after the mercury resistance gene regulator of *E. coli*. The members of this family are involved in stress-response and detoxification gene regulation in bacteria. ZntR (Outten *et al.* 1999), CueR (Outten *et al.* 2000), PMTR (Noll *et al.* 1998) and MerR (Summers 1992) bind metal ions and activate heavy-metal resistance mechanisms. SoxR responds to oxidative stress through iron-sulfur centers (Hidalgo *et al.* 1997b). NolA is involved in the nodulation process in *Bradyrhizobium japonicum* by responding to nodulation factors from soybeans (Sadowsky *et al.* 1991). BmrR (Ahmed *et al.* 1994) and TipAL (Holmes *et al.* 1993) bind and are activated by drugs. Presumably Mta and BltR also bind drugs, although their natural ligands are still unknown. The MerR family is defined by sequence homology in the first ~110 residues of the proteins, but the C-terminal domains vary significantly, displaying only limited examples of homology even within the family (Figure 1.10).

The majority of early work on the family was concentrated on MerR and its ability to preferentially bind mercury over other metal ions. MerR binds DNA in the middle section of a promoter that controls transcription of mercury detoxification proteins and overlaps with its own promoter (Heltzel *et al.* 1987). It represses transcription of the detoxification genes in the absence of mercury, activates them in the presence of Hg(II), and always represses it own transcription (Heltzel *et al.* 1987). It is able to repress transcription in the absence of mercury, even while bound to an inactive RNAP, in a

mechanism called "active repression" (Heltzel *et al.* 1990). There are three conserved cysteine residues (residues 82, 117 and 126), which are necessary for the binding of metal ions (Zeng *et al.* 1998). These cysteine residues are located on opposite sides of a predicted antiparallel coiled-coil (Lupas 1997; Wolf *et al.* 1997; Zeng *et al.* 1998) and bind Hg(II) with a trigonal coordination site using C82 from one monomer and C117 and C126 from the other (Shewchuk *et al.* 1989b; Helmann *et al.* 1990; Ralston *et al.* 1990; Utschig *et al.* 1995). Upon binding of a single mercury ion per dimer, a conformational shift takes place which underwinds the DNA and allows open promoter formation (Ansari *et al.* 1992; Summers 1992; Ansari *et al.* 1995). Mutational studies have been done which identify these and other residues as being important for signaling, metal binding, DNA binding and activation by MerR (Shewchuk *et al.* 1989a; Shewchuk *et al.* 1989b; Parkhill *et al.* 1993; Comess *et al.* 1994; Parkhill *et al.* 1998).

The protein members of the MerR metal-binding subfamily tend to be short, with lengths of 135 residues in *E. coli* CueR to 144 residues in *E. coli* MerR, and to share homology throughout the proteins. The C-terminal domains of the proteins of the drug-binding subfamily, on the other hand, tend to be significantly larger. While the metal-binding subfamily only extends between 25 and 40 residues beyond the conserved domain, the drug-binding C-terminal domains average around 140 residues. This correlates with the size and complexity of the ligands. A few well-placed cysteine residues are sufficient to specifically bind a metal ion, however in order to recognize and bind the larger ligands, it appears that a larger protein domain is required. TipAL is able to bind covalently a 1.4 kD multicyclic peptide antibiotic, thiostrepton, but doing so

requires a C-terminal domain of 144 residues (Holmes *et al.* 1993). BmrR binds a variety of cationic lipophilic molecules with its 158 residue C-terminal domain (Markham *et al.* 1996).

Even before the structure of any MerR family member had been solved, these proteins were known to contain at least two structurally competent domains. TipAL has an alternative translation start site that results in a protein, TipAS, which is only the C-terminal domain of TipAL beginning with Met 110 and running through the end of the protein at residue 253 (Holmes *et al.* 1993). TipAS is found to bind the same large antibiotic as TipAL and is thought to be a mechanism for modulating activity of TipAL and sequestering the antibiotic (Chiu *et al.* 1996). The C-terminal half of BmrR, called BRC for BmrR C-terminus, was crystallized and its structure solved both with and without bound drug (Zheleznova *et al.* 1999) indicating that this domain alone retains the ability to fold and bind drugs specifically. NolA is translated in three different lengths from three alternative translation initiation sites (Loh *et al.* 1999). $NolA_1$ contains the full-length protein, while $NolA_2$ and $NolA_3$ are N-terminally truncated and lack the HTH motif. However, only the N-terminal 109 amino acids exist in MtaN, a constitutively active mutant form of Mta from *B. subtilis* (Baranova *et al.* 1999). Thus these proteins have at least two independently folded structural domains, one corresponding to roughly the first 110 residues, and the remainder of the proteins comprising the other domain(s).

A set of MerR family member proteins is involved in the regulation of multidrug transporters in *B. subtilis*. As previously described, the *bmr* (bacterial multidrug

resistance) and *blt* (Bmr-like transporter) genes both encode major facilitator superfamily (MFS) pumps capable of transporting and conferring resistance to a similar spectrum of compounds (Ahmed *et al.* 1995). Regulator genes encoding MerR family members BmrR and BltR, are found near the *bmr* and *blt* genes, respectively. Transcription of *blt* has not been detected in wild-type cells and is not inducible by any of the known Blt substrates. Activators of BltR are unknown, however it has also been shown to control the other gene in the operon, *bltD*, which encodes an acetyltranferase involved in spermidine degradation (Ahmed *et al.* 1995; Woolridge *et al.* 1999). This suggests that the natural role of the Blt transporter may be in efflux of polyamine metabolites. The *bmr* gene, on the other hand, is inducible by many of the substrates of Bmr through BmrR. BmrR binds as a dimer to the imperfect inverted repeat spacer between the –10 and –35 boxes in the *bmr* promoter These consensus hexamers, the -10 sequence and the -35 sequence, are critical for binding and open complex formation by RNA polymerase. BmrR represses transcription until it binds a drug, at which time the *bmr* gene is activated. A DNase I footprint of BmrR on the *bmr* promoter covered all of the spacer region, most of the –35 box and some of the –10 box, approximately 29 bp in all (Ahmed *et al.* 1994). When BmrR binds an activator molecule, it alters the DNA in the promoter to allow productive interaction with RNAP. A third MerR family member, Mta, appears to regulate both *bmr* and *blt* genes (Baranova *et al.* 1999).

Discovery of MtaN

Mta (multidrug transporter activation) was discovered in screens for MDR mutants of *B. subtilis* by Baranova, *et al.* in 1999 (Baranova *et al.* 1999). The group was

looking for new mutations in *bmr* or *blt* that would lead to an MDR phenotype. They plated *B. subtilis* on LB (Luria-Bertani) agar containing 10 μgrams ml^{-1} of ethidium bromide. Ethidium bromide is a known mutagen, bacteriostatic agent and substrate of the Bmr and Blt transporters. Bmr overexpression through BmrR action was found in 22 of 25 resistant clones. In the other three clones, *mta1*, *mta2* and *mta3*, resistance was due to overexpression of Bmr and Blt but was independent of BmrR. Genomic DNA from these clones was included on plasmids and found to confer MDR upon wild-type strains and upon those lacking either Bmr or Blt, although at reduced effectiveness. Northern blot analysis confirmed that both Bmr and Blt were overexpressed in strains with the *mta1* mutation, including cells that lacked BmrR. These results suggested that the mutation was in a transcriptional activator capable of activating both *bmr* and *blt* independently of BmrR.

The mutation was mapped by phage transduction and PCR to a hypothetical gene called *ywnD*, which was renamed *mta* (multidrug transporter activation). Sequencing revealed a single nucleotide substitution which created a stop codon after codon 109 of the gene. The two other clones mentioned above, *mta2* and *mta3*, were also found to be the results of mutations in this region. Those mutants had frameshifts that led to truncation of the Mta protein at nearly the same position as *mta1*. The mutant protein was termed MtaN for Mta, N-terminus.

Truncation of the *mta* gene could give rise to MDR by two possible mechanisms. The first, in which truncation of the gene led to loss of function and thereby gain of MDR,

was discounted by the finding that complete deletion of the *mta* gene did not lead to MDR. The second possibility was that the truncated Mta protein is a constitutive activator of the *bmr* and *blt* genes. To confirm this hypothesis, *mta1* and full-length *mta* were cloned into plasmids and transformed into cells. The cells receiving *mta1*, which encodes MtaN, gained the MDR phenotype by overexpressing Bmr and Blt, while those receiving *mta* did not (Baranova *et al.* 1999).

Sequence analysis of this gene revealed a MerR family member, most closely related to the TipAL protein of *Streptomyces lividans* (Holmes *et al.* 1993), but also with significant N-terminal homology to BmrR and BltR. The finding that a truncated form of Mta, containing the DNA-binding and dimerization domains but lacking the C-terminal ligand-binding domain, is a constitutive activator led to the conclusion that the C-terminal domain inhibits transcriptional activation by these proteins. Binding of ligand, or removal of the domain, relieves this inhibition and allows activation of the genes under the control of the protein.

MtaN was shown, through electrophoretic-mobility shift assay and DNase I footprinting, to bind to the *bmr* and *blt* promoters, and its own promoter. Binding of full-length Mta was only proven to occur at the *mta* promoter by these methods. However, the K_d of MtaN for DNA was not determined. MtaN is a constitutive activator and autogenously regulated, leading to a positive feedback loop that creates high levels of MtaN in *mta1* cells. By Northern blotting, the authors were able to show that MtaN activated transcription of a hypothetical membrane protein, YdfK. YdfK is found

downstream of its own hypothetical MerR family gene, *ydfL*. Thus, the authors showed that MtaN could bind and/or activate transcription at three dissimilar MerR family promoters. However, whether this protein could bind and activate all MerR promoter was left unanswered.

Mechanism of activation

MerR family member proteins bind and activate a promoter structure with an unusual 19 base pair (bp) spacer between the –10 and –35 boxes (Parkhill *et al.* 1990), which places these recognition elements in an orientation different from the average bacterial promoter, which has a 17 bp spacer. The most widely accepted model for transcriptional activation by these proteins is that the binding of a single coactivator molecule causes a conformational change in the dimer that leads to an undertwisting and shortening of the effective length of the DNA. The conformational changes in DNA were thought to bring the key promoter elements an orientation that would allow for open promoter complex formation (Ansari *et al.* 1992; Summers 1992; Outten *et al.* 1999). This theory was formulated based upon the work done on MerR and its cognate promoter. Whether this is a common mechanism across the family remains to be determined.

This mechanism was addressed when the first crystal structure of a MerR family member was reported. BmrR was solved bound to DNA and co-activator (Zheleznova-Heldwein *et al.* 2001), which is relevant to MtaN for several reasons. First, both are MerR family member proteins from *B. subtilis*, and contain the signature N-terminal DNA-binding domain of the family. They share 28% sequence identity through the length of MtaN. Moreover, both proteins are known to be able to activate transcription at the

bmr promoter, despite having several different residues on the recognition helix. Because of their similarities, it was predicted that they will have similar structures and mechanisms of transcriptional activation, but whether their mechanisms are identical is unknown.

BmrR structure

BmrR was solved by a combination of molecular replacement, using the previously solved BRC (BmrR C-terminus, figure 1.11) (Zheleznova *et al.* 1999), and multiple isomorphous replacement (MIR, described in the following chapter). The refined structure contained a ternary complex of BmrR, DNA, and drug (TPP^+, tetraphenylphosphonium), presumably in a transcriptionally active conformation. The overall structure of the dimer was described by the authors as resembling "a butterfly perched on a bent stick" (Figure 1.12). The protein contains three domains: an N-terminal four-helix bundle, which contains a winged-helix DNA binding motif; an 11-turn dimerization α-helix; and a C-terminal drug-binding domain. The bound DNA is bent, untwisted and crumpled, thus shortening its effective length. A bound TPP^+ ion is found in the same binding pocket as that observed in the BRC structure.

The dimerization of BmrR buries ~5,800 $Å^2$, mainly through interactions of the drug binding domain of one monomer with the DNA-binding domain of the other, which bury ~2,000 $Å^2$ per monomer. The antiparallel coiled-coil formed by the previously mentioned 11-turn α-helices also buries a significant amount of surface area. This antiparallel coiled-coil has been predicted to occur in all MerR family members (Caguiat *et al.* 1999), and forms the main dimerization interface found in MtaN, as it has no "drug-binding" domain. The winged-helix motifs are found on the same side of the BmrR

dimer, with their recognition helices nearly parallel to each other separated by ~30 Å (Figure 1.12). These winged-helix motifs contain the typical HTH, followed by a short β-sheet (wing 1) and a second pair of helices in an orientation similar to that of the HTH. Between these helices is a short turn, which has been called wing 2 because it also contacts the DNA (Zheleznova-Heldwein *et al.* 2001).

Even though the complex displays a significant number of interactions between the protein and DNA, base-specific contacts were not seen. The 3.0 Å resolution of the BmrR-DNA-drug complex may be insufficient to visualize the base-specific contacts, although one would expect any such contacts to be fairly well-ordered and visible at this resolution. Alternatively, there may be no base-specific contacts in the complex, with all specificity being achieved through indirect readout, on the basis of the DNA structure rather than specific base contacts. This is unlikely because the promoters recognized by the MerR family are likely to have similar DNA structure, making it difficult for a protein to discriminate between them on that basis. The inability to model base-specific contacts is more likely due to the crystal form of the complex, which formed with only a half-site in the asymmetric unit. Because the *bmr* promoter is an imperfect palindrome, the bases and specific protein-DNA interactions will differ on either side of the dimer. Since the dimer is related by crystallographic symmetry, the density at any point will be an average of the two possible orientations, causing statistical disorder when the monomers (or half-sites) differ from each other. This statistical disorder will obscure the correct density. However, since MerR family members are believed to be always bound to DNA and only activated upon binding of ligand (Ahmed *et al.* 1994), it is also possible that the base-

specific contacts are made only in the BmrR-DNA binary complex, and upon activation the specific contacts are dissolved as the complex changes conformation. Confirmation of this awaits the structure determination of the BmrR/*bmr* binary complex.

The DNA structure of the complex reveals a novel mode of activation. As noted previously, the MerR family promoters contain an unusually long 19 bp spacer between the -10 and –35 elements, placing them in an improper orientation for productive binding by RNAP. In the BmrR/DNA/drug complex, the DNA is significantly distorted from B-form. The DNA is bent approximately 50° away from the protein. DNA bending is common among DNA-binding proteins, MarA (Rhee *et al.* 1998) and PurR (Schumacher *et al.* 1994) being but two examples. The DNA in the center of the complex contains unpaired bases and is crumpled. This distortion leads to a shortening and undertwisting of the DNA, as had been predicted, which brings the –10 and –35 elements into an orientation more closely resembling that found in promoters with 17 bp spacers between them. The crumpling of the DNA means the consecutive major grooves of the DNA match the ~30 Å distance between the two recognition helices of the protein. However, the short distance between recognition helices and consecutive major grooves is insufficient to explain the undertwisting of the DNA. Many DNA-bending proteins are found to bend DNA by compressing the major grooves without untwisting the DNA. Undertwisting the DNA indicates some other type of motion on the part of the protein.

Another question left unanswered in the absence of the BmrR-DNA binary complex is the signaling of the drug binding to the dimer partner. On possibility is that

helix α6 of the drug binding domain reorganizes upon drug binding, causing a shift in the position of the DNA-binding domain. The other possibility is that the drug-binding signal is transmitted through the coiled-coil dimerization domain to the other subunit (Figure 1.12). The latter is supported by mutational studies on MerR (Comess *et al.* 1994; Caguiat *et al.* 1999), although both possibilities could certainly be true. This question also awaits structural information on the BmrR/*bmr* binary complex.

Conclusion

Structural studies on MtaN were undertaken for several reasons. MtaN is a constitutive activator, unlike other MerR family members studied to date. While the DNA-binding and activation properties differ among other MerR family members, presumably in part because of the presence of C-terminal ligand-binding domains, MtaN lacks this section of the full-length molecule. Thus, the structure of MtaN will represent a form of a MerR family member unencumbered by the influence of this domain. Since the ligand-binding domain is believed to hold the protein in a transcriptionally inactive form until binding of ligand, it would be interesting to determine what form the protein would take without this domain. Does MtaN take an active or inactive formation off the DNA and without the presence of the C-terminal domain?

The ability of MtaN to activate other MerR promoters led to the conclusion that it was a global MerR activator. That it activated the *bmr* and *blt* promoters led to the naming of the protein as the multidrug transporter activator, although a degree of binding specificity for MDR promoters over other MerR family promoters was not shown. The ability of MtaN to up-regulate *ydfK* supported the conclusion that it was a global

activator, and to the suggestion that YdfK might be an MDR transporter. Structural and binding studies would assist the understanding of the mechanism of binding by MtaN to various promoters. Is MtaN truly a global activator? What specificity, if any, does it display? How does MtaN recognize, bind and activate its own and other promoters?

At the time this work began, no structure of the DNA-binding portion of a MerR family member had been solved. Before the structure of MtaN was solved, the structure of BmrR in complex with DNA and a drug was completed. However, MtaN remained a promising and important target. Because of its uncommonly high solubility in the absence of DNA or drug ligands, MtaN could be crystallized in the apo form, which had not been reported for any MerR family protein. The structure of the apo form of a MerR protein would allow examination of changes taking place in the protein upon binding of DNA. Could the apo form of the molecule already be in a conformation suitable to binding of B-form DNA, or the form of DNA found in the BmrR complex? Does the BmrR model of DNA-activation apply for MtaN and other members of the MerR family?

Not all of these questions are answered in this thesis. The results described here do significantly illuminate the form of apo MtaN and the ability of MtaN to bind and activate multiple MerR promoters. This work has allowed us to propose a structural model with some new features for the mechanism of DNA activation by MerR family proteins. A more complete understanding of the binding and activation mechanisms by this family of proteins awaits the structure of MtaN in complex with DNA.

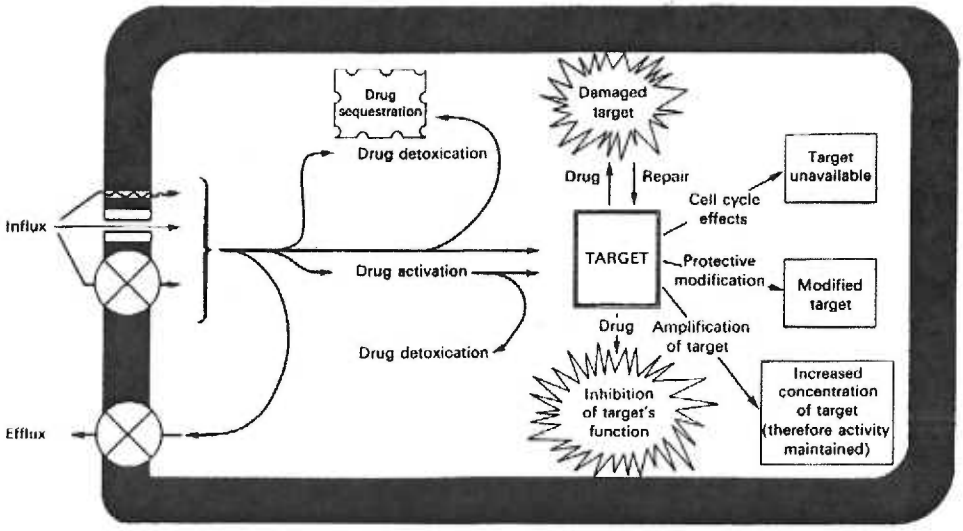

Figure 1.1 Biochemical mechanisms of drug resistance. From (Hayes *et al.* 1990).

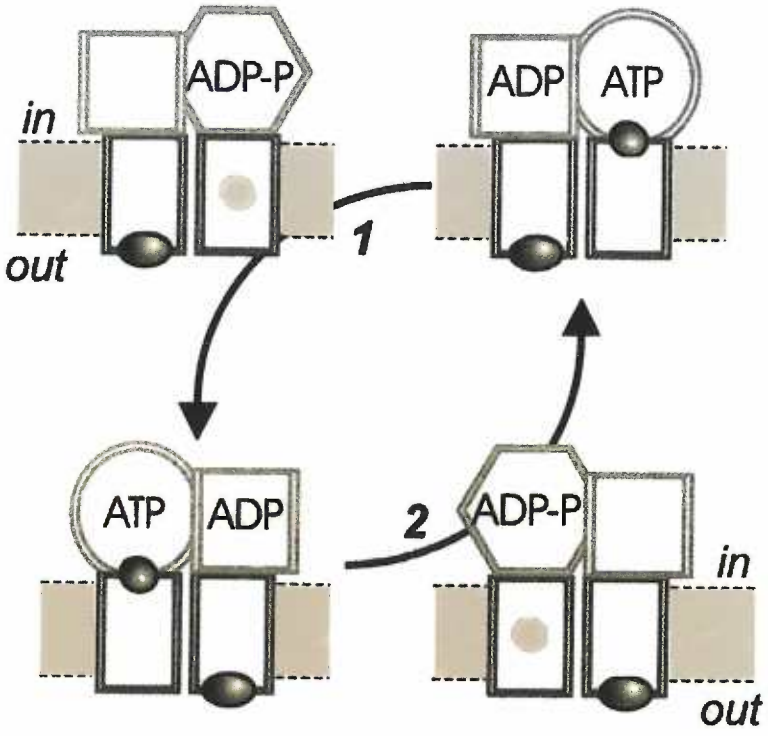

Figure 1.2 A two-site model for the action of ABC multidrug transporter LmrA. The circles, squares and hexagons represent different nucleotide-bound forms of the cytoplasmic nucleotide-biding domains (NBD). In the presence of ATP the cycle follows the arrows. (i) The ATP-bound NBD has a high-affinity drug site on the interior of the cell, while the ADP-bound has a low affinity site on the outside of the cell (upper right). (ii) When the ATP is hydrolysed, the drug moves toward a newly created low-affinity site because the high-affinity site is disrupted, and the drug is released from the low-affinity site on the partner, creating an empty ATP binding site (upper left). (iii) ATP and drug bind to the dimer partner as the previous drug is moved to the outside of the cell (lower left), and the cycle is repeated. From (van Veen 2001).

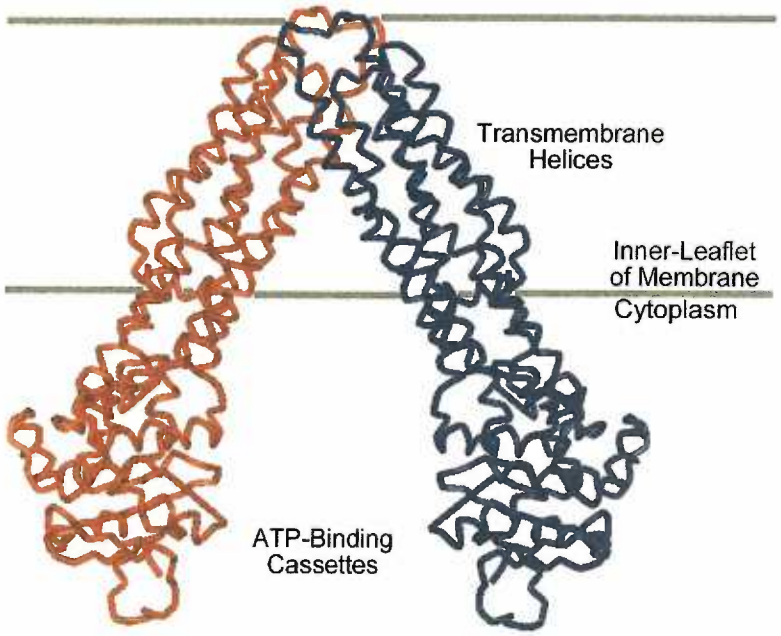

Figure 1.3 The 4.5 Å resolution crystal structure of MsbA from *E. coli*, a homolog of the ABC transporters. Due to low resolution, only the α-carbon trace is shown. Near the inner leaflet of the membrane is an opening to the space between the dimer partners. This opening is believed to be the entrance for drug binding.

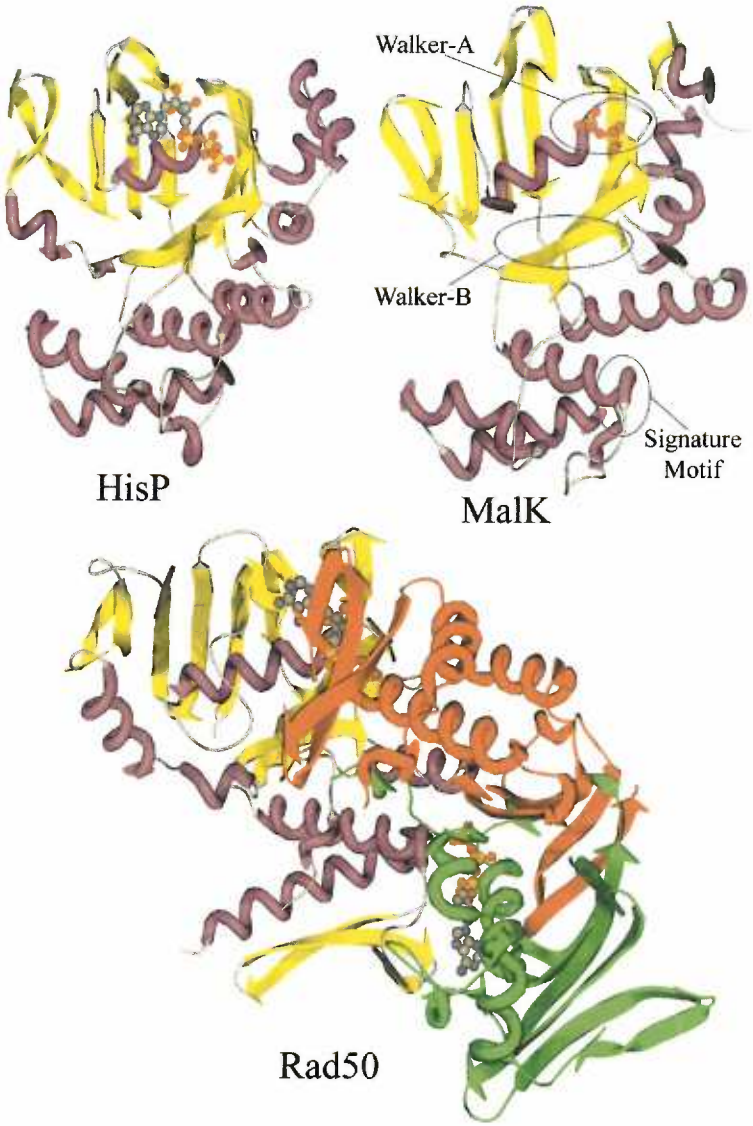

Figure 1.4 Models of three ABC domains showing the locations of the conserved Walker A and B and signature motifs. Bound nucleotide can be seen in the Rad50 and HisP models. MalK was solved without nucleotide.

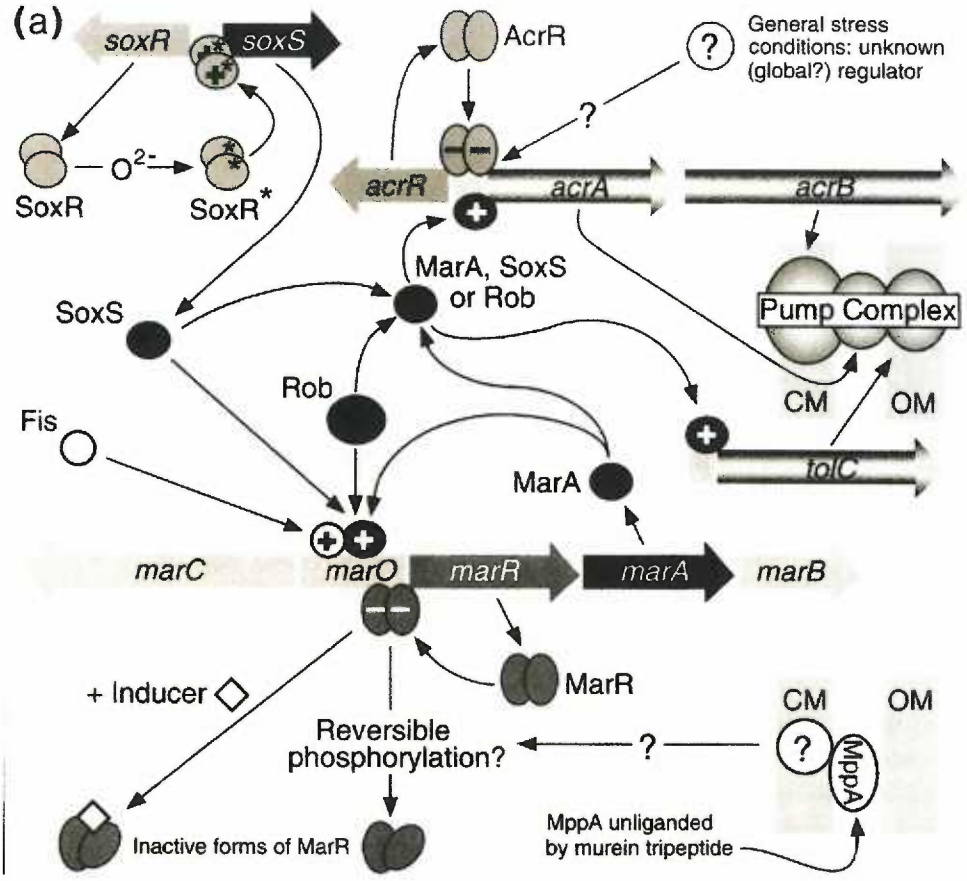

Figure 1.5 A schematic representation of MDR regulation in *E. coli* from (Grkovic *et al.* 2001b). While MarA is the centerpiece of the regulation scheme, it does not bind inducers and is a global constitutive activator. Inducer response occurs by MarR binding drugs and regulating the level of MarA produced.

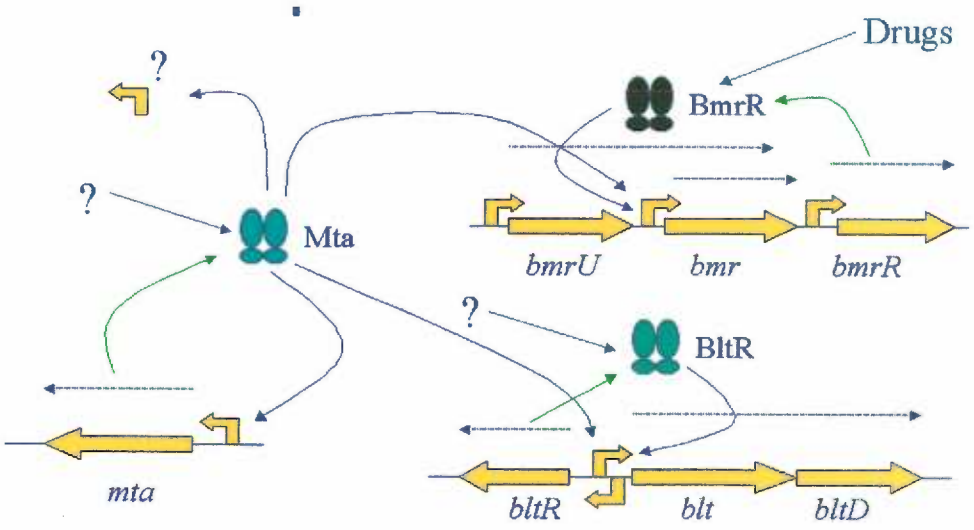

Figure 1.6 Model for transcriptional control of the MDR transporter genes *bmr* and *blt* of *B. subtilis*. BmrR and BltR regulate their own genes and those of their cognate transporters, and both are further controlled by Mta. Inducers of BmrR include cationic lipophilic molecules, while inducers of Mta and BltR remain unknown.

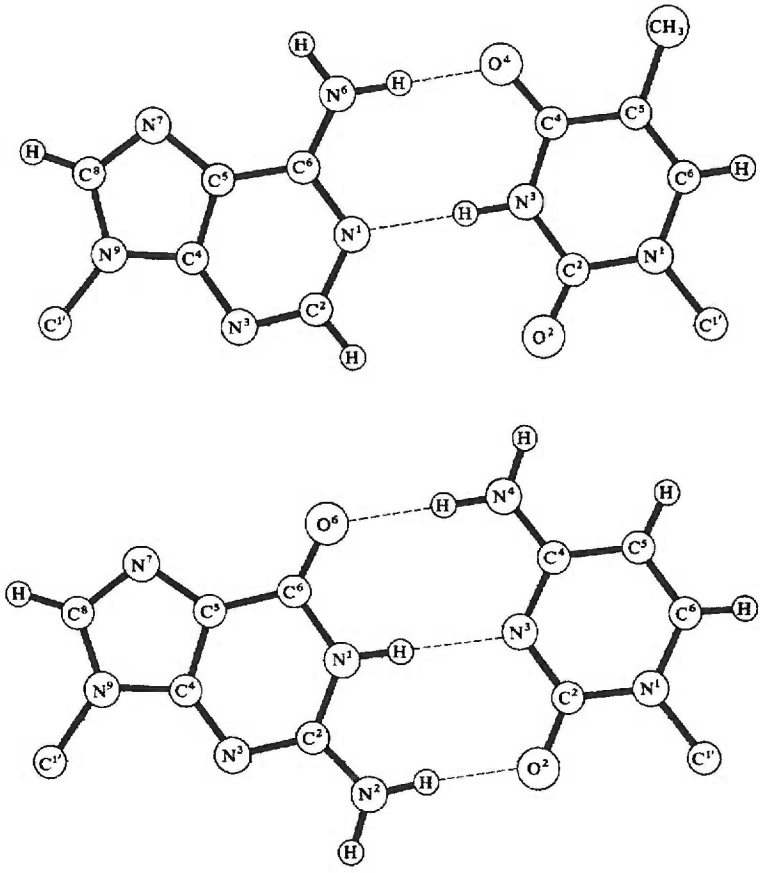

Figure 1.7 A-T (top) and G-C (bottom) base pairs as found in B-DNA. The major groove is towards the top of the page and the minor groove is towards the bottom of the page. Hydrogen bonds are shown as dotted lines. Atom names are shown. Hydrogen bonding potentials of the atoms in the major and minor grooves are discussed in the text. Adapted from (Cantor *et al.* 1980), page 178.

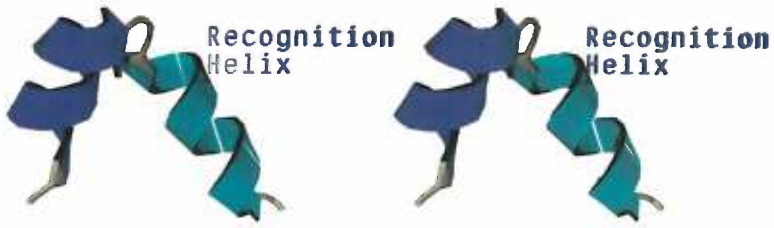

Figure 1.8 A stereo view of the HTH motif from MtaN.

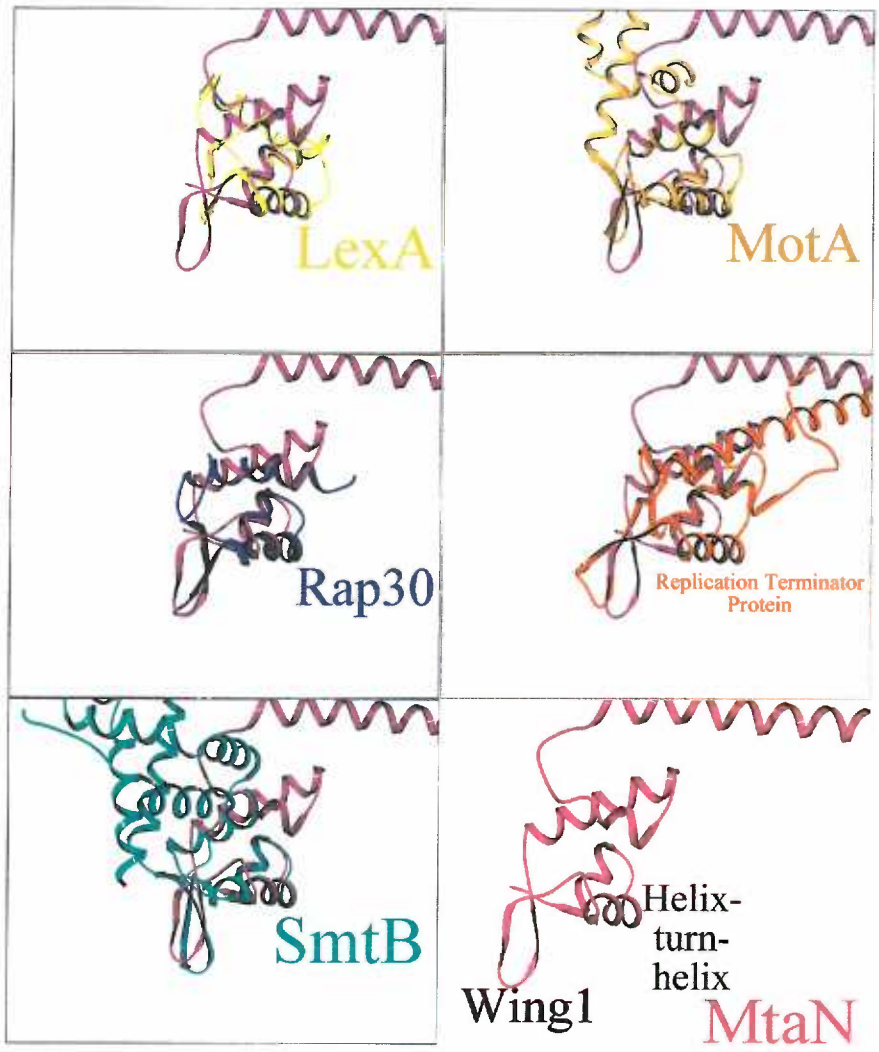

Figure 1.9 Overlays of some winged-helix proteins of known structure onto the structure of MtaN. MtaN is in purple in all frames. Only the HTH sections of the proteins were used to calculate the overlays. Despite significant diversity, all contain a loop region after the recognition helix. These loop regions are of various structures and functions. Structures shown are taken from the Protein Data Bank: LexA (1LEA), MotA (1BJA), Rap30 (1BBY), Replication Terminator Protein (1BM9) and SmtB (1SMT).

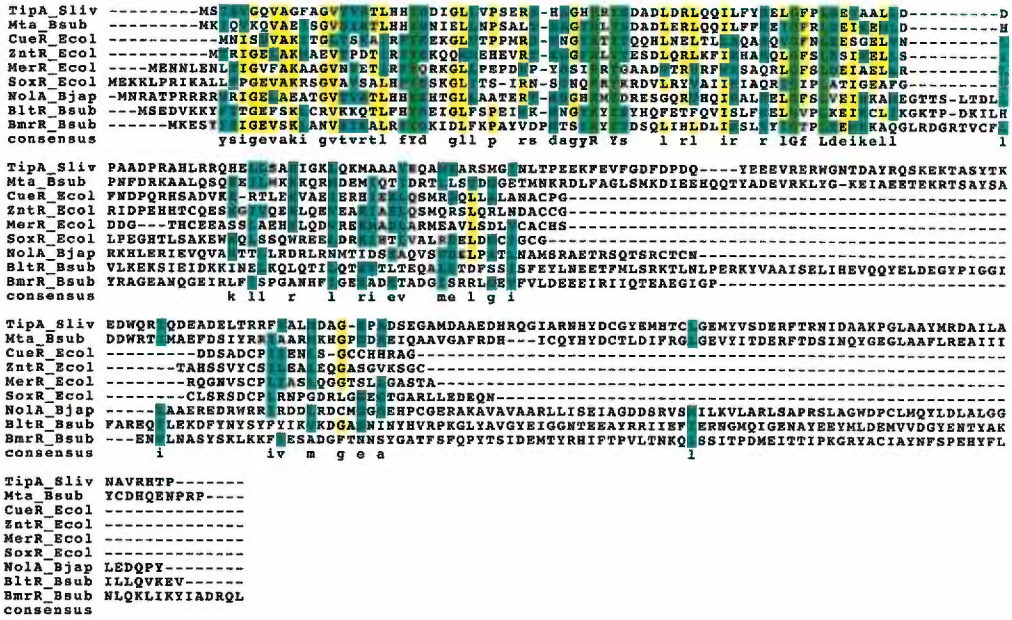

Figure 1.10 An alignment of nine of the known MerR family members. Most of the homology seen across the family is in the first 110 residues.

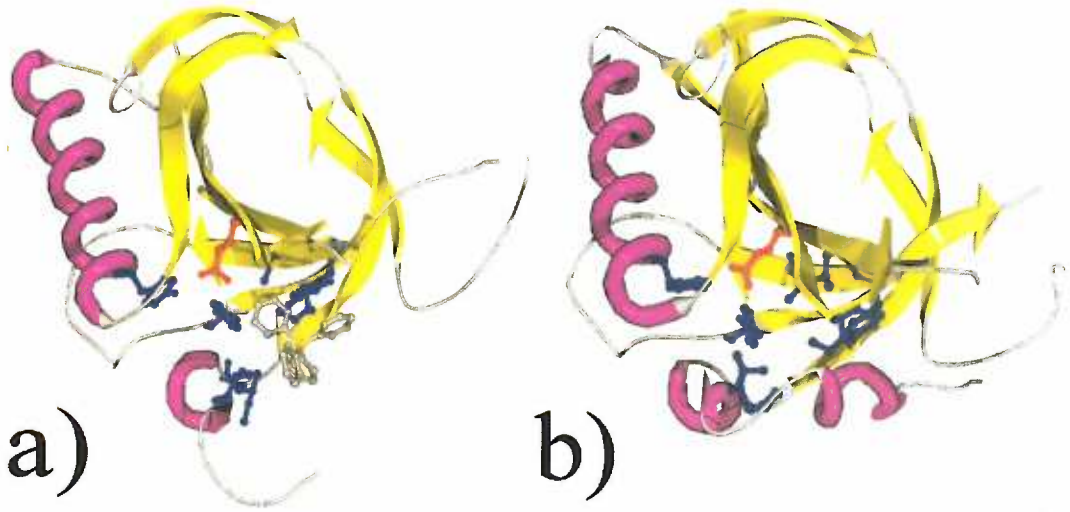

Figure 1.11 The drug-binding domain of BmrR, called BRC. A buried glutamate is in red, and hydrophobic side chains found to contact TPP^+ are in blue. a) BRC bound to tetraphenylphosphonium. b) BRC without bound drug. A short α-helical shield covers the drug-binding site and protects it from solvent. This helix is disordered once the drug binds.

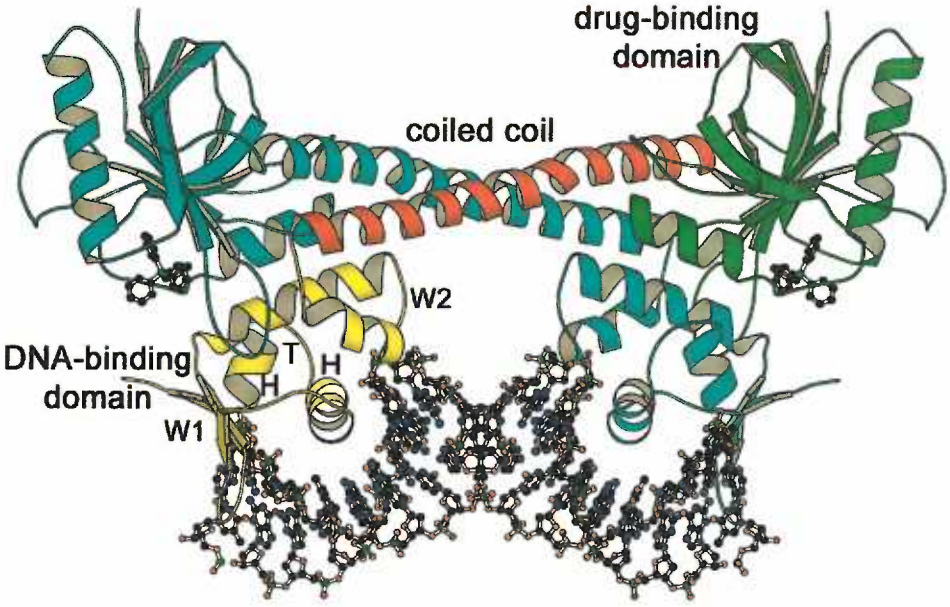

Figure 1.12 A ribbon representation of the BmrR dimer bound to DNA and TPP^+. One monomer is color coded yellow (DNA-contacting domain), red (dimerization antiparallel coiled-coil helix) and green (drug-binding domain). The other is colored blue. The structure shows a large bend in and effective shortening of the bound DNA. From (Heldwein *et al.* 2001).

Chapter 2

MACROMOLECULAR CRYSTALLOGRAPHY

Introduction

As early as 1912, protein crystals had been shown to diffract X-rays. Somewhat ironically, this was the first solid evidence that proteins contained a regular, ordered structure down to the atomic level. If proteins had been merely long chains of amino acyl residues, without defined 3-dimensional structures, they would not have crystallized and their crystals would not have diffracted X-rays, and there would be no protein X-ray crystallography. The first protein structure solved by X-ray crystallography was myoglobin in the late 1950s. Since then, the number of solved structures has increased to tens of thousands, and the rate at which "new" structures are being released is still increasing with the advent of newer and more powerful techniques and computers.

A simplistic explanation of macromolecular crystallography is summarized here. Protein X-ray crystallography is a process whereby X-ray diffraction patterns, from crystals grown of purified protein, are measured and converted to three-dimensional maps of electron density. Into these electron density maps, molecules are placed piece by piece until a model of an entire protein is built, which matches the calculated electron density. An X-ray crystal structure determination proceeds through a series of regular steps (Figure 2.1). First, a protein of interest is overexpressed and purified. Crystals are grown through a process of trial-and-error using a variety of crystallization solutions. When crystals of suitable volume have been grown, they are placed in an X-ray beam and

diffraction measurements are made (and the crystal characteristics are determined; unit cell, space group, etc.). These measurements consist of the intensities of thousands of diffracted spots. Phases are then calculated (or estimated) for each of these spots. With phase and intensity information, electron density maps are calculated. A protein model is then built to fit into the calculated electron density, and this model is refined against both the measured data and what is known about general protein structure. The final refined model is then studied carefully to relate structure to function.

Practical crystal growth

Purified protein is concentrated to several milligrams per milliliter, and this protein is then used for crystallization trials. One crystallization method of choice is the hanging-drop diffusion method (McPherson 1999). In each well of a 24-well plate, a 1 mL quantity of a crystallization solution is placed. In order to find initial crystallization conditions, a variety of crystallization solutions are sampled initially. Two μL of a crystallization solution is mixed with two μL of the protein solution on a glass coverslip. This coverslip holding the protein-crystallization solution mix is inverted and placed over the well containing the crystallization solution, and sealed with grease to form a self-contained microenvironment. In this microenvironment, the drop contains approximately one-half of the concentrations of crystallization reagents as the well below it. This situation is energetically unfavorable. In order to reach an equilibrium state, the water from the drop will slowly transfer to the well through diffusion, raising the concentrations of the crystallization reagents in the drop, thereby slowly decreasing the solubility of the protein. As the protein solubility decreases, it will reach a state of supersaturation and

come out of solution (Figure 2.2). If the solubility of the protein decreases at a slow enough rate, and to enough of an extent, the protein may come out of solution in the form of a crystal or crystals. Crystallization will also decrease the amount of protein in solution, keeping the protein concentration on the edge of supersaturation. If the change in solubility is too rapid, or the protein reacts poorly to the crystallization solution by denaturing or aggregating, the protein may precipitate out in an amorphous solid form. On the other hand, if the solubility of the protein does not decrease significantly enough, it will stay in solution. Some variables that can be tested in searching for suitable crystallization conditions include: pH, protein concentration, temperature, salts, detergents, and precipitant (chemical nature or concentration).

Crystals and Crystallization

Crystals are composed of regularly repeating units of molecules. These units are called unit cells, and contain one or more molecules of interest. The unit cell is described by variables **a**, **b**, and **c** (the lengths of the sides) and α, β, and γ (the angles between the sides). Angle α is found between sides **b** and **c**, while angle β is between **a** and **c**, and so on (Figure 2.3). In many crystals, these variables turn out to be simple or repeated values; in space groups other than triclinic (described below), the most common angles are 90° and 120°. Furthermore, in many crystals **a** may equal **b** or **c**, or both. Combinations of these relationships between variables leads to a variety of regular ways in which crystals can grow, called crystal systems.

Crystal systems are ways of further describing crystal architecture and symmetry. There are only seven crystal systems: monoclinic, triclinic, tetragonal, cubic,

orthorhombic, trigonal, and hexagonal. Triclinic crystals may assume any values for the six variables listed. Monoclinic crystals have **a** **b** **c**, $\alpha = \beta = 90°$, and $\gamma > 90°$, thus the number of independent variables is decreased from 6 to 4. If $\alpha = \beta = \gamma = 90°$, the cell can be orthorhombic (**a** **b** **c**), tetragonal (**a** = **b** **c**), or cubic (**a** = **b** = **c**), reducing the number of independent variables to 3, 2, or 1 respectively. Hexagonal cells have $\alpha = \beta = 90°$, $\gamma = 120°$, and **a** = **b** **c**. Trigonal cells are characterized by $\alpha = \beta = 90°$, $\gamma = 120°$, and **a** = **b** = **c** in a hexagonal lattice; **or** $\alpha = \beta = \gamma \neq 90°$, and **a** = **b** = **c** in a rhombohedral lattice, depending on the axes used. These crystal systems describe the shape and size of the unit cells.

Within these crystal systems, there are several subtypes of arrangements. Lattice points can always be found at the corners of the unit cell, such as in the primitive lattice, signified by the letter P. There can also be other lattice points in the unit cell, leading to non-primitive cell types. A system with a lattice point in the center of the unit cell is called body-centered, denoted by I (Inner-centered), and is exemplified by MtaN crystals, the topic of this thesis. If lattice points exist in the middle of all the faces, it is called face-centered (F). Lastly, if extra lattice points are found on only one set of opposite faces, it is called A, B, or C-centered for the faces on which the lattice points are centered. A cell with lattice point in the center of the C side is C-centered. The primitive lattices plus the non-primitive lattices are the 14 Bravais lattices (Figure 2.4).

Within any lattice of symmetry higher than triclinic, that is to say all of the other crystal systems, the arrangement of objects in the cell is related by symmetry operations.

Each of these objects can be described as an asymmetric unit. An asymmetric unit may be a molecule, a dimer of molecules, a complex containing several molecules, or some other arrangement of molecules. However, the asymmetric unit is repeated in the unit cell through regular symmetry operations. Thus, each asymmetric unit in the unit cell (and in the entire crystal) is identical except for orientation and position. Only one asymmetric unit needs to be "solved" to solve the internal structure of the entire crystal. In the case of MtaN, there is one molecule per asymmetric unit, and eight asymmetric units per unit cell. The dimer of MtaN is related by crystallographic symmetry, so that the structure of only one monomer must be solved to determine the structure of the dimer.

Crystallographic symmetry is created by symmetry operations. There are several types of symmetry operators which can relate one asymmetric unit to another, including mirror planes, rotations, glide planes, inversions, and screw axes. Combinations of these operators can transform one asymmetric unit to all the others in the unit cell. These combinations are called point groups and there are 32 possible point groups. When the 32 point groups are combined with the 14 Bravais lattices, mathematicians found that there are exactly 230 possible ways for crystals to grow, called space groups. Proteins have internal asymmetry because all of their amino acids are enantiomeric (L-type). Thus, symmetry operations involving mirrors or inversion centers are inaccessible to protein crystals. This fact reduces the number of space groups available to protein crystals to only 65. They are denoted by a capital letter for the type of lattice (Primitive, Inner-centered, Face-centered, A, B, or C face-centered), and numbers signifying the internal symmetry operations. For example, MtaN crystals grow in the very rarely taken space

group $I2_12_12_1$, meaning body-centered with three perpendicular 2_1 screw axes (Figure 2.5). (The orthorhombic geometry of the MtaN unit cell is deduced from the symmetry elements.)

X-ray diffraction

X-rays are a form of electromagnetic wave like visible light but with much higher energy. The energy of a wave increases as the wavelength (λ) decreases; visible light has λ from 4,000 to 7,000 Å (Å = 1 x 10^{-10} meters), while X-rays have λ from 0.1 to 100 Å. This is the form of energy used in macromolecular crystallography because its λ includes atom-atom bond lengths (~1.5 Å for carbon-carbon bonds). X-rays used for in-house experiments are produced by electronic bombardment of a copper target and have a wavelength of 1.54 Å, but synchrotron radiation also often used, which is of a shorter λ and higher intensity than can be achieved in a small laboratory setting. Synchrotron radiation is a by-product of the constant acceleration of a charged particle moving around a circular path (Figure 2.6). Synchrotron radiation can be tuned to different λ, which is vital for the phase determination technique used on the MtaN crystal structure (Leigh 1976).

As early as 1912, Max Von Laue had observed the diffraction of X-rays by crystals. His contemporary W. L. Bragg described the diffraction spots as "reflections" because they appeared to be like reflections off planes occurring within the crystal lattice. These lattice planes are called the *hkl* planes and can be shown to cross the unit cell an integral number of times (Figure 2.7). The number of parallel planes that intersect each

side of the unit cell will determine its *hkl* designation. If a set of parallel planes cross the **a** side once, the **b** side twice and the **c** side three times, those are the 123 planes. While the spots are not truly reflections off of these planes, they behave similarly, thus they are still often referred to as reflections. Bragg's geometric derivation of the properties of these planes produced the law named for him:

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

This law governs the diffraction patterns produced by crystals and relates spots to the angle θ of the spots and the wavelength λ of the radiation, to the perpendicular ( ) d spacing of the planes that produce those spots. The d spacing is a measure of how far apart the parallel planes, described above, are from each other. The reason for Bragg's law is fairly simple, when shown in diagrammatic form (Figure 2.8). Given that X-rays are transverse waves reflecting off parallel planes, then one can imagine the interference of these reflected waves would only be constructive at certain points. In order for the waves to interfere constructively after one wave has traveled farther than another, the difference in distance that the two traveled would have to be a multiple of the wavelength. In this case, the wavelength is λ, and the extra distance traveled by the wave to get to the distant parallel plane and back will equal twice the distance between those planes (the d spacing), multiplied by the sine of the angle of the beam to those parallel planes. Only when this relation is satisfied will interference be constructive and a diffraction spot appear. This leads to the relationship that sin θ is proportional to the inverse of the d spacing. An inverse relationship is less satisfactory than a direct one, so in order to

simplify the use of the d spacing to describe diffraction, a reciprocal lattice has been devised. The reciprocal lattice is not based on d , but on 1/d . This construction also requires the use of reciprocal axes and angles between the reciprocal axes, instead of the real axes. In orthogonal crystal systems, such as that found in MtaN crystals, the real axes and the reciprocal axes point in the same directions, such that $\alpha = \alpha^*$, but are of different lengths and units, so that a and a* (reciprocal a) are related by $a^* = 1/a$ (Figure 2.9). Reciprocal axes are measured in 1/Å instead of regular Å. Construction of a reciprocal lattice is straightforward and extremely helpful in understanding diffraction (Figure 2.10). Construction of a reciprocal lattice begins with the real lattice. From a lattice point designated as the origin (O), a line normal to a real-space reflecting (lattice) plane is drawn as described above. This line will be of a length inverse to the closest distance from the plane to O. The point at end of this line will be named for the plane from which it came. This is the reciprocal lattice point representing that plane. Then points are drawn for planes in all directions. These points make up the reciprocal lattice. Each reciprocal lattice point represents one reflecting set of planes.

To understand the value of the reciprocal lattice, the *sphere of reflection* is used to describe reflections in more visual terms. A 2-dimensional reciprocal lattice can be constructed with a lattice point O at the origin (Figure 2.11). A circle of radius $1/\lambda$ is drawn, which has O on the edge, point C in the center and X as the source of the X-rays, directly opposite O. Thus ray XCO goes from the X-ray source, through the center of the circle C and point O, passing through the reciprocal lattice of the crystal along the way. The circle is the 2-dimensional representation of the sphere of reflection. Any place

where the circle touches a reciprocal lattice point (points P and P' on the diagram) will produce a diffraction spot. The intensity of this spot depends upon the electron density on the *hkl* plane from which it originates. Proof of this follows.

One more point B is placed on the circle (Figure 2.11), opposite from O and on line XO. Triangle BOP will then be a right triangle since it is inscribed in a semicircle. Define angle PBO as θ. Because the radius of the circle is 1/λ:

$$\sin\theta = {OP}/{BO} = {OP}/{2/\lambda}$$

Therefore:

$$\sin\theta \; {2}/{OP} = \lambda$$

Because of the way the reciprocal lattice was constructed, the length of OP is equal to $1/d_{hkl}$ where *hkl* are the identifiers for the reflecting planes from which that reciprocal lattice point came. Thus $1 / OP = d_{hkl}$. Substitution into the previous equation yields:

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

This is Bragg's law for $n = 1$. Furthermore, when the crystal rotates around O, the reciprocal lattice rotates with the real lattice such that whenever a reciprocal lattice point crosses the circle of reflection, Bragg's law is satisfied and a reflection occurs. This construction will also predict the directions in which reflections will travel.

The direction of reflections can be calculated given that, during the construction of the reciprocal lattice, the direction of any OP_{hkl} is normal to the reflecting plane *hkl*. Because BOP is a right triangle, BP is normal to OP and thus is parallel to the original reflecting plane. Given a line parallel to BP but going through C, the angle at which the X-ray approaches this line is θ, thus the reflected ray is bent by an angle of 2θ from the incident beam and passes directly through P_{hkl}. This reflection is named after the *hkl* planes from which it came. By rotating the crystal, all the reciprocal lattice points within the limiting sphere can be made to contact the sphere of reflection at least once, and thus potentially produce reflections. The limiting sphere has a radius of $2/\lambda$ about the origin and represents all possible reflections from a crystal, although most of the reflections will not be unique. The reflections *hkl* and *-h -k -l* can be viewed as being the same reflection off the opposite sides of the *hkl* plane. These reflections are a Friedel pair and are "equivalent" reflections. In the absence of anomalous diffraction, the intensities measured from these spots should be the same. This is Friedel's law:

$$I_{hkl} = I_{\overline{hkl}}$$

Friedel pairs are not the only examples of equivalent reflections. Space groups other than triclinic have internal symmetry which creates other sets of equivalent reflections. Thus, there are symmetry elements not only in the crystal, but also in the diffraction pattern (Laue symmetry). Laue symmetry creates a number of equivalent reflections, which will depend upon the space group. In the orthorhombic space group $I2_12_12_1$, taken by MtaN, the Laue symmetry is 222 and:

$$I_{hkl} \equiv I_{\bar{h}\bar{k}l} \equiv I_{\bar{h}kl} \equiv I_{h\bar{k}l} \equiv I_{hk\bar{l}} \equiv I_{\bar{h}\bar{k}\bar{l}} \equiv I_{\bar{h}k\bar{l}} \equiv I_{h\bar{k}\bar{l}}$$

thus, only 1/8 of the possible reflections will be unique. Because of equivalent reflections, it is possible to collect complete data sets for high-symmetry crystals with only a partial rotation of the crystal. Internal crystal symmetry will also create missing reflections, for which the measured intensity should always be zero. These are extremely helpful in space group determination for a new crystal form.

Of course, when study of a new crystal begins, the unit cell dimensions, the crystal system, and the space group are unknown until diffraction data are collected. By measuring the angles of many reflections and understanding the reciprocal lattice and the sphere of reflection, one can work backwards from the diffraction pattern to derive **a**, **b**, **c**, and α, β, and γ. In most cases the space group can be unambiguously determined from the pattern of reflections, including missing and equivalent reflections. Space groups have characteristic patterns of reflections that are either zero or equal due to symmetry operations within the unit cell. This information is only the overall structure of the unit cell, and does not reveal the arrangement of electrons within the unit cell. That information comes from interpretation of the structure factor amplitudes ($|F|$), which are calculated from the measured intensities (I) of the diffraction spots by:

$$I_{(h\,k\,l)} = |F_{(h\,k\,l)}|^2$$

with some corrections for diffraction properties, which are described later. The electron density along each of the planes of reflection will affect the intensity of the diffracted

spots, thus there is a relationship between the arrangement of electrons in the unit cell and the intensities of the diffracted spots. Knowing the geometry and precise arrangement of the contents of the unit cell, it is a straightforward process to predict the positions and relative intensities of diffraction spots. However, the reverse is not nearly as simple because of the "phase problem" that will be discussed later. One can measure the positions and relative intensities of diffraction spots and then work backwards to find the 3-dimensional electron density pattern that would create the relative intensities observed in the diffraction pattern, only if one has the phase of each reflection.

Structure factor equation

Each diffracted X-ray has a wave nature, and is therefore periodic. Hence, X-rays can be described by a Fourier series, which is called the structure factor equation. Because each atom in the unit cell contributes to the each diffraction spot, each spot contains information from each atom. The equation relating the structure factor, $\mathbf{F}_{hkl}$, to the contribution from each atom can be written as:

$$\mathbf{F}_{hkl} = \Sigma_j f_j \exp^{2\pi i (hx_j + ky_j + lz_j)}$$

where f_j is the scattering factor from the j^{th} atom, i is the imaginary square root of -1 (used to identify the imaginary component of scattering), x_j, y_j and z_j are the fractional coordinates of the j^{th} atom, and hkl are the previously described indices of reflection. Thus, every atom contributes to every reflection in an amount determined by its scattering factor (f) and its position in the unit cell ($hx + ky + lz$). At the end of this chapter, a thermal parameter (B) is described which also affects the atomic scattering factor.

However since X-rays are scattered by electrons, which are not merely points in space, the equation is also written as a function of electron density ρ:

$$\mathbf{F}_{hkl} = \int_x \int_y \int_z \rho(x, y, z) \exp^{2\pi i(hx + ky + lz)} dxdydz$$

or in terms of volume V:

$$\mathbf{F}_{hkl} = \int_V \rho(x, y, z) \exp^{2\pi i(hx + ky + lz)} dV$$

thus $\mathbf{F}_{hkl}$ is a Fourier transform of the electron density in the unit cell. Because a Fourier transform is a reversible operation, the measurable $|F_{hkl}|$ can be converted to the function of interest, $\rho(x,y,z)$ by representing it as a Fourier transform of the structure factors:

$$\rho(x,y,z) = (1/V) \Sigma_h \Sigma_k \Sigma_l |F_{hkl}| \exp^{-2\pi i(hx+ ky + lz)+i\alpha_{hkl}}$$

in which V is the unit cell volume, and $|F_{hkl}|$ is the structure factor amplitude, which can be obtained from measurements of reflection intensities as previously described ($I(hkl) = |F(hkl)|^2$). The final variable in the equation is α_{hkl}, which is the phase of each reflection. While this is required for calculation of $\rho(x,y,z)$ it cannot be measured, which gives rise to the "phase" problem. It must be determined through the use of phasing techniques, which will be described later.

X-ray data collection

X-rays for our studies are generated in one of two ways, either by bombarding a copper target with electrons, or by synchrotron radiation. In either case, the produced X-rays are directed at the protein crystal, which has been mounted either in a small capillary

or a nylon loop. The crystal diffracts these rays, for the reasons described above, into a pattern of discreet spots. A detector is placed behind the crystal to collect information from the diffraction pattern. Historically these detectors used photographic film, but they have been modernized, mostly with charge-coupled device (CCD) or electronic image plate detectors.

In order to measure individual reflections, it is necessary to take a series of "pictures" of the diffraction pattern. Each picture, or frame, is a time-lapse exposure of a small rotation of the crystal, typically between 0.5° and 1.5° depending on the density of reflections. If the diffraction pattern is very dense, a large rotation will lead to several unrelated spots overlapping simultaneously upon the detector surface. Overlapping spots cannot be used, in general, because there is no satisfactory method for determining how much radiation came from each reflection. However, if the oscillation angle is too small, too many reflections will be spread out over more than one frame. It is possible to reconstruct these "partial" reflections, but if they are too numerous, the current statistical methods used to combine them become less precise, and data quality suffers.

To measure the intensity of the spots, an appropriate-sized box is drawn around each discreet spot. The darkness from each pixel in the box is added together to produce the measured intensity for that spot. In other words, the intensity of the spot is integrated over the area of the box by the computer. This process is done for every spot on a frame. Then the next frame is loaded and the process repeats. To normalize the intensity in each frame to the other frames, given the assumption that over enough spots the additive

intensity on each frame should be similar, each frame is assigned a scale factor, which multiplies the intensity of every spot on the frame. This is followed by the processing of the partial reflections, and each reflection's intensity is output to a file. All this is done to collect, integrate, scale, and merge intensities data for each crystal. To measure the internal agreement of a data set, the intensities of "equivalent" reflections are compared.

The goal of any intensities data collection strategy is to collect 100% of all reflections to a given resolution. However, a data set containing 90% of the accessible measurements is usually adequate, and often the completeness of the data set will be significantly less in the higher-resolution shells. Many equivalent reflections will be measured in the course of regular data collection, and these will provide a measure of the reliability and accuracy of the data collection process. Equivalent reflections measured during data collection never have exactly equal intensities, mainly because of detector error, integration error, and crystal damage caused by X-ray exposure. The differences between "equivalent" symmetry-related reflections provide a measure called the R_{sym}:

$$R_{sym} = \frac{\sum_{hkl}\sum_{i}\left|I_{i(hkl)} - \overline{I_{(hkl)}}\right|}{\sum_{hkl}\sum_{i} I_{i(hkl)}}$$

where $\overline{I_{(hkl)}}$ is the average intensity of the i observations of a reflection. A lower R_{sym} is more desirable because it signals a higher level of agreement within the data set. Less than 10% is generally considered acceptable, although 7% or less is better.

Phase determination

When a data set has been collected and processed to $|F|$s, there is still not enough information to calculate electron density maps and build a model. As mentioned, in the section on the structure factor, the structure factor amplitudes are not enough. Phase information for each reflection is required in order to calculate the 3-dimensional electron density, $\rho(x,y,z)$.

One way to determine phases for a data set is to use molecular replacement (MR) (Rossmann 1990). Molecular replacement uses a known structure from a protein which is likely to have a similar structure as a starting model to find phases for a data set. Molecular replacement works because a known structure can be "moved" around within the unit cell by the computer to determine where in the unit cell the known structure best matches the measured data. EPMR (evolutionary program for molecular replacement) is a computer program which uses an evolutionary algorithm to find a molecular replacement solution (Kissinger *et al.* 1999). It works by generating a population of search molecules with random orientations and locations within the unit cell. A correlation coefficient is calculated for each molecule, which relates it to the observed structure factor amplitudes. A fraction of the highest-scoring molecules are used to create a new population set for the next round by applying random shifts to the "surviving" solutions. This process repeats for a number of rounds, or until a threshold level of correlation coefficient is reached. At the end, a regular rigid-body refinement (described later) is done to fine-tune the position and orientation of the best molecule. Other MR programs separate the rotational and translational searches to find the correct solution,

because a brute-force search of trying every possible rotation/translation combination is impractical. Finding a rotation solution first, then doing a translation search with that solution significantly reduces computing time. When successful, the output of any MR program is a correctly placed search molecule, no matter what strategy was used. The "molecular replacement solution" is used to estimate phases and is then refined and rebuilt to match the experimental contents of the unit cell. This can be done when a protein of known structure is grown in a crystal with a different space group, or when a protein of unknown structure is similar to a protein of known structure. Using the structure of MerR family member BmrR (Zheleznova-Heldwein *et al.* 2001), molecular replacement phasing failed in the structure determination of MtaN. Although surprising, this result indicated that BmrR was not structurally similar enough to use as a search model. When MR is not possible, the phase problem must be solved by other methods.

De novo phasing methods: MIR and MAD

To determine a *de novo* structure, one must use more elaborate methods of structure determination. The theory behind these methods relies on the knowledge that every atom in the structure contributes to each diffraction spot, and that different atoms diffract X-rays more or less strongly. Heavier atoms have more electrons around them and diffract X-rays more strongly than lighter atoms. The most common elements in protein structures are carbon, nitrogen and oxygen, which have six, seven and eight electrons, respectively. These are all fairly light elements and diffract weakly. Furthermore, they have electrons in only two energy levels. These two properties of protein atoms allow the *de novo* phasing techniques described below.

The first method of *de novo* phasing is multiple isomorphous replacement (MIR) (Ke 1997). MIR takes advantage of the fact that heavier elements, e.g. metals, diffract X-rays more strongly. A set of protein crystals must be grown, and some of the crystals are soaked in solutions containing heavy atom compounds, such as platinates, mercurials, or silver or lead salts. Alternatively, protein crystals can be grown in the presence of these compounds. The goal is to bind heavy atom compounds to specific sites on the protein at 100% occupancy. For example, Hg^{2+} often specifically reacts with the sulfur atom of cysteine residues. The intensities diffraction data of these "derivatized" proteins can be used to find the phases needed to calculate electron density maps. Unfortunately, the intensity data of many derivative crystals cannot be utilized because their geometry is significantly different from the native crystals. This is referred to as nonisomorphism, and precludes the multiple isomorphous replacement method of phase determination.

An alternative is the highly successful and now commonly used multiwavelength anomalous diffraction (MAD) method (Hendrickson 1991). MAD uses only one crystal for the "derivative" and "native" data sets, and thus there is perfect isomorphism. However, these crystals must contain an anomalously scattering atom with an absorption edge in the tunable range of X-ray wavelengths for synchrotron radiation. While sulfur contains an absorption edge in the tunable range and is normally found in proteins, heavier atoms such as selenium or bromine give stronger anomalous signals. To generate a protein that contains a strongly anomalously scattering atom, the protein under study must be grown in the presence of selenomethionine and absence of regular methionine. By doing so and shutting down the metabolic pathways of the expressing cells that would

normally make methionine, either through genetic alteration or metabolic repression (Van Duyne *et al.* 1993), the bacteria will incorporate selenomethionine into all protein positions that would normally contain methionine. Fortunately, the selenomethionine-substituted protein crystals tend to grow in conditions similar to the native crystals, with similar properties, although the native crystals often diffract better. Hendrickson was able to show that even one selenium atom in a small protein of about 150 residues is enough for structure determination by this method (Hendrickson *et al.* 1990). This method also has the convenience of having "heavy atoms" at known locations in the protein sequence, a fact which aids model building significantly. Because an anomalously scattering atom will have different contributions to the overall scattering at different wavelengths, it will act as a variable in an otherwise "constant" diffraction pattern. By varying the wavelength of incident radiation around the absorption edge of selenium, and collecting intensity data sets on different points of the absorbance curve, selenium gives information similar to that of a heavy atom in MIR. By locating the selenium atoms, one can use their phase information to generate electron density maps.

Using MAD techniques to solve the structure of MtaN followed this set of steps:

1. Grow protein crystal in which methionine is substituted by Se-methionine.

2. Perform XAS (X-ray absorption spectrometry) scan on the crystal to confirm presence of Se and find absorbance edge.

3. Collect intensity data sets which are at the inflection point and the peak of the edge, and at a remote wavelength to serve as the "native" set, and convert intensity data to structure factor amplitudes, $|F|$.

4. Use Patterson difference methods to find positions of anomalously scattering atoms.

5. Calculate phases and electron density.

6. Improve maps and build model.

XAS is a method for finding the specific absorption edge of the anomalous scatterers in the crystal, because absorption edge energies are slightly dependent upon the chemical environment of the atom. In order to find the specific edges for an anomalous scatterer in a protein, the crystal is mounted in line with the synchrotron beam and a fluorescence monitor is placed perpendicular to the beam. The wavelength of the beam is scanned around the predicted absorption energy, while the monitor measures the fluorescence of the crystal, which is compared to the intensity of the primary beam. The trace of the ratio of fluorescence to beam intensity against wavelength (energy) will give a clear indication of the wavelength (and energy) and shape of the absorption edge.

To carry out the intensities data collection in a MAD experiment, two wavelengths are chosen at specific points on the absorption edge and at least one remote wavelength for a "native" data set to optimize signals from the selenium atom. By choosing at least three wavelengths, both the dispersive and Bijvoet differences can be

maximized. f and f are the real and imaginary components, respectively, of the anomalous scattering atoms. One point chosen is at the peak of absorption, which is also the peak of f . Another is at the inflection point of the absorption spectrum to maximize the difference in f , which maximizes dispersive differences. At least one wavelength remote from the edge should be collected for comparison (Figure 2.12). The f and f contribute to the total scattering, **F**:

$$\mathbf{F} = \mathbf{F}_P + f_h + f' + f''$$

where $\mathbf{F}_P$ is the sum of scattering from all the atoms without anomalous scattering (the light atoms), and f_h is the normal scattering from the "heavy" atom. This will be described in more detail later.

At each of these wavelengths a MAD data collection scheme is implemented to make paired measurements of intensity from "equivalent" reflections as close together in time as possible. Friedel pairs are reflections from opposite *hkl* indices (I_{hkl} and I_{-h-k-l}). Reflections are a Bijvoet pair if an odd number of sign changes are required to transform the *hkl* indices and if they related by a symmetry element. Thus, in an orthorhombic space group, *h*, *k*, *l* and -*h*, *k*, *l* are a Bijvoet pair while, *h*, *k*, *l* and -*h*, -*k*, *l* are not. The Friedel mate of a reflection is always a Bijvoet. In the absence of anomalous scattering, intensities of Friedel or Bijvoet pairs are equivalent, however, they display differences due to anomalous scattering. Because these Friedel mates display a Bijvoet difference in MAD data, and these differences must be measured as precisely as possible, data are collected from an angle range ϕ_1 to ϕ_2, and then from $\phi_1+\pi$ to $\phi_2+\pi$; this is referred to as

the "inverse beam" method (Hendrickson 1991). This strategy gives reflection sets of $\{h,k,l\}$ and $\{-h,-k,-l\}$ in close succession. The same angles, and thus the same set of $\{h,k,l\}$ and $\{-h,-k,-l\}$ reflections, for each of the wavelengths are collected consecutively to ensure that the dispersive (λ-dependent) differences are also measured as accurately as possible. This is important because in a typical MAD experiment the differences in scattering factors are <10 electrons, much smaller than the ~ 80 electrons of a heavy atom such as mercury. Even though intensity data collection on cryo-cooled crystals significantly reduces X-ray induced crystal damage, this general scheme is still followed to minimize any effects on X-ray data collection because the crystal will be exposed to high-power radiation for a long period of time and some damage will inevitably occur (Weik *et al.* 2000).

Polarization and Lorentz corrections are applied to the measured intensities. A polarization correction is applied because there is a dependence of scattering efficiency on the polarization of the incident radiation and its relationship to the scattering angle, and synchrotron radiation is highly polarized (Drenth 1999). A Lorentz correction accounts for how fast the reflection is passed through the sphere of reflection; those near the axis pass through more slowly, sometimes so slowly that the Lorentz correction is unreasonably high and they must be excluded. Once the MAD data are collected, they are processed and integrated essentially the same as any other data (described above) except that multiple data sets are produced from one collection, and that the Friedel pairs are kept separate to preserve the Bijvoet differences. If collected correctly, these data are sufficient to calculate phases and electron density maps, but to accomplish this the

anomalous scatterers or "heavy atoms" must be located. This is done by the use of difference Patterson maps, just as in MIR.

Patterson synthesis

Patterson maps can be used in order to find the relative positions of heavy atoms, either the metal atoms of the derivative, or the anomalously scattering atoms in MAD. The Patterson function:

$$P(u, v, w) = 1/V \, \Sigma_h \Sigma_k \Sigma_l \, |F_{\mathrm{PH}(hkl)}|^2 \, \mathbf{cos}[2\pi(hu + kv + lw)]$$

or in exponential form:

$$P(u, v, w) = 1/V \; \Sigma_h \Sigma_k \Sigma_l \, |F_{\mathrm{PH}(hkl)}|^2 \, \exp^{-2\pi i(hu + kv + lw)}$$

where u, v, and w are fractional coordinates in Patterson space, is phase-independent and uses the measurable square of the structure factor amplitudes, which is proportional to the measured intensities. However, unlike the x, y and z coordinates of an electron density map, the u, v and w coordinates of Patterson peaks are not real-space coordinates, but interatomic coordinates (Figure 2.13). The height of a Patterson map peak is proportional to the product of the number of electrons in the two atoms that create the peak. Thus the highest peaks on a Patterson map will be from vectors between the heaviest atoms, or in the case of MAD as MIR phasing, between atoms with the largest anomalous scattering differences. This is important because a Patterson map will contain a huge number of peaks; each vector between two atoms will create a peak, such that the number of Patterson peaks will equal $n^2 - n$, where n is the number of atoms in the unit cell. Even in

a small protein like MtaN, which contained 872 atoms in the final structure and 8 molecules per unit cell, this relationship would result in over 45×10^6 Patterson peaks. Most peaks will be proportional to between 36 and 64 electrons^2, but the peak from the vector between two mercury atoms, for example, will be proportional to 80^2 or 6,400 electrons^2. A further gain in signal-to-noise is made because difference Pattersons are used, where F_{PH} in the above equations is replaced by $(|F_{PH}| - |F_P|)^2$ giving the difference Patterson equation:

$$\Delta P(u, v, w) = 1/V \, \Sigma_h \Sigma_k \Sigma_l \, ||F_{PH(hkl)}| - |F_{P(hkl)}||^2 \exp^{-2\pi i(hu + kv + lw)}$$

By using this coefficient, the contributions of the native atoms to the Patterson function will be diminished and the derivative peaks will stand out. However, even using difference Patterson maps, it can sometimes be difficult to locate heavy atom coordinates.

Patterson peaks fall into predictable patterns. By using the symmetry operators of the space group we can predict where some of the heavy-atom Patterson peaks be located in one dimension at a time. In the $I2_12_12_1$ space group, an atom at x, y, z, will be matched by a symmetry related atom at $\frac{1}{2}+x$, $\frac{1}{2}+y$, $\frac{1}{2}+z$, so there will be an interatomic vector and corresponding Patterson peak at ½, ½, ½ for each atom in the asymmetric unit. The other operators mean that there will be also be interatomic vectors on the planes 0, v, w; u, 0, w; u, v, 0; ½, v, w; u, ½, w; and u, v, ½. These "Harker sections" are used to find and cross-validate possible heavy-atom Patterson peaks. Real peaks appearing on one Harker section should be found at related locations on the other Harker sections. Once the heavy-atom sites are located, their positions are converted to real-space and refined, and phases

are calculated. In the case of MAD data, anomalous scattering properties are also refined slightly to more closely match the data.

In the case of MtaN, the selenium atoms were near each other in space, and close to symmetry elements. This made the difference Patterson maps difficult to interpret visually. Use of the computer program SOLVE (Terwilliger *et al.* 1999) aided the search for selenium sites by analyzing a large number of possible Patterson peaks. SOLVE calculates difference Patterson maps, finds peaks in the Patterson density and scores potential heavy-atom solutions. Potential solutions are refined then scored on the basis of agreement with the Patterson, difference Fouriers, the presence of "solvent" and "protein" regions in a native electron density map, and the figure of merit. The best potential solutions are then used in self-difference Fouriers to search for other possible "heavy atom" sites. Three of five potential sites were located in this way. The fourth and final site was located using the real-space difference Fourier function:

$$\Delta\rho(x,y,z)=1/V\ \Sigma_h\Sigma_k\Sigma_l\ |F_{hkl(\text{deriv})}-F_{hkl(\text{native})}|^2\ \exp^{[-2\pi i(hu+kv+lw)+i\alpha_{hkl}]}$$

where α_{hkl} are the phases calculated from the initial selenium sites. A large peak in the difference Fourier map revealed the location of the fourth selenium site. The fifth possible site, on the N-terminal selenomethionine, was never found and was either disordered or not present due to normal protein processing by *E. coli*.

Calculating phases based on MAD data

The structure factor equation includes the phase α of each reflection (Figure 2.14). However, as discussed, α is not directly measurable. In MIR, the measurable quantities

are $|F_P|$ and $|F_{PH}|$, the structure factor amplitudes from the protein alone and the heavy atom derivatized proteins, respectively. It is also possible, based upon the determined locations of the heavy atoms, to calculate $|F_H|$ and α_H, the structure factor amplitude and phase of the heavy atoms. These quantities can be related by:

$$|F_{PH}| \exp(i\alpha_{PH}) = |F_P| \exp(i\alpha_P) + |F_H| \exp(i\alpha_H)$$

so that the unknown quantities in the equation are α_P and α_{PH}, and the one required in order to calculate electron density maps is α_P. This is an ideal case, as measurements will always contain errors. Treatment of these errors is discussed later.

It is not immediately obvious how to obtain the missing α_P. The phase can be found through the use of a geometric construction proposed by Harker in 1956 (Harker 1956). The principle is simple, yet elegant. On a coordinate with one real and one imaginary axis, first draw a circle about the origin with radius $|F_P|$, because $|F_P|$ is measurable (Figure 2.15). The vector representing $\mathbf{F}_P$ must be somewhere on this circle, because the circle contains all possible phases. Because $|F_H|$ and α_H have been calculated (although they still contain errors which will be discussed later), one can then draw the vector $-\mathbf{F}_H$ from the origin (since $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$, then $\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H$). $|F_{PH}|$ has also been measured, so it is possible to draw a circle of radius $|F_{PH}|$ centered on the end of the $-\mathbf{F}_H$ vector. Since $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$, one of the vectors to the two intersections of the circles must be $\mathbf{F}_P$ and contain the correct α_P. Unfortunately, they are equally probable, so another heavy atom derivative must be collected to resolve the phase ambiguity.

The Harker construction works with similar principles in MAD phasing, but is done using anomalous signals instead of heavy atom signals. As previously described, in the case of MAD, the total scattering factor (**F**) is:

$$\mathbf{F} = \mathbf{F}_P + f_h + f' + f''$$

where $\mathbf{F}_P$ is the sum of scattering from all the light atoms, and f_h is the normal scattering and f' and f'' are the real and imaginary parts of the anomalous scattering from the "heavy" atom. Because of the physical properties of anomalous scattering, f' is 180° out of phase and f'' is 90° out of phase with f_h (Figure 2.16). The total scattering vector **F** will be dependent upon wavelength due to the dispersive (wavelength-dependent) differences in f' and f'', while $\mathbf{F}_P$ and f_h remain essentially constant around the absorbance edge. These changes are mathematically analogous to isomorphous derivative changes. Again, only $|F|$ can be measured (determined from the intensities), but by varying the wavelength to optimize the signals from f' and f'', the differences in $|F|$ and α_{PH} are significant enough to use the Harker construction.

Construction of the Harker phase solution for MAD as MIR begins the same as with MIR. A circle of radius $|F_P|$ is drawn around the origin of a coordinate. In this case $|F_P|$ is the measurement at a "remote" wavelength not immediately on the absorption edge; the measurements at the edge will be labeled $|F_{PH1}|$ and $|F_{PH2}|$.

$$\mathbf{F}_{H1} = \mathbf{F}_{PH1} - \mathbf{F}_P$$

is used to calculate $\mathbf{F}_{H1}$. The vector $\mathbf{F}_{H1}$ is drawn from the origin. Then a circle of radius $|F_{PH1}|$ is drawn around the head of the $\mathbf{F}_{H1}$ vector. Where the circles intersect are the possibilities for the phase α_P of that reflection. The ambiguity can be solved by drawing the $\mathbf{F}_{H2}$ vector from the origin and a circle of radius $|F_{PH2}|$ around the head of that vector. An alternative to this is to recall that because the imaginary component of the anomalous scattering is 90° out of phase with f_h, that the Friedel pairs will not be equal. The difference between them will be due entirely to the flip of f . Thus $\alpha_{P(hkl)}$ can also be solved using the information from I_{hkl} and I_{-h-k-l} (Figure 2.17). In practice, both methods are used simultaneously to overdetermine the phases. Once all the $\alpha_{P(hkl)}$ have been determined and $|F_P|$ measured, an electron density map can be calculated.

In practice, there are always errors in measurements and phase angle determinations. While these errors are minimized by the use of a single crystal and data collection run, they are never completely eliminated. Thus phases come with a measure of uncertainty, and how this uncertainty is treated will affect the quality of the resulting electron density map. Briefly, measurement errors manifest themselves in the Patterson maps, then the determined heavy atom positions and anomalous scattering contributions, thus the calculated phases, and therefore in the electron density maps. Attempts to minimize or compensate for error are made at every stage. Measurements of related reflections are collected to minimize X-ray damage effects to the crystal by collecting them close in time. Patterson maps are made using difference maps by subtracting the native Patterson from the "heavy" Patterson. Initial heavy atom positions are refined against the intensity data to more accurately describe their locations and scattering

properties. However, the most complete treatment of errors is done once phases are determined.

In the discussion of the Harker construction, perfect circles were drawn because it was assumed that all $|F|$ measurements were accurate and precise. In practice, this is not the case and thus the circles are not perfect lines, rather they have density and width to them corresponding to the uncertainty of their of measurements. Thus, the intersections of these diffuse circles are not discreet points, but are more accurately represented by probability curves (Figure 2.18). More accurate measurements give sharper curves and better phases.

Blow and Crick (Blow *et al.* 1959) developed the "lack of closure" method of dealing with errors (Figure 2.19). It requires that phases have already been calculated. This method assumes that $\mathbf{F}_{H(calc)}$ and $\mathbf{F}_{P(calc)}$ have been determined correctly and that all error is in the measured magnitude of $\mathbf{F}_{PH}$. The $\mathbf{F}_{H(calc)}$ and $\mathbf{F}_{P(calc)}$ vectors are added and $\mathbf{F}_{PH}$ is pointed towards the resultant vector, but will nearly always be, as observed, too long or too short. This allows the definition of the lack of closure error (ε) at a given phase angle as:

$$\varepsilon(\alpha_P) = |F_{PH}|_{obs} - |F_{PH}|_{calc}$$

where $|F_{PH}|_{calc}$ is determined from $\mathbf{F}_{H(calc)} + \mathbf{F}_{P(calc)} = \mathbf{F}_{PH}$ as described previously. This allows the calculation of a phase probability curve by calculating $P(\alpha)$ over all possible angles:

$$P(\alpha) = \exp(-\varepsilon(\alpha)^2 / 2E^2)$$

where E is the estimate of error from:

$$\langle E^2 \rangle = \langle(|F_{PH} \pm F_P| - |F_H|)\rangle.$$

With a single derivative, the curve will have two equally tall peaks, representing the two possible phase angles from the Harker construction. So, the product of probability curves from different derivatives is used to resolve the ambiguity between the peaks. Hendrickson and Lattman formulated an alternative, which stores the probability function in four coefficients to make it easier to use.

Only in a probability curve with a monomodal distribution will the highest point directly correspond to the best phase. By choosing the highest point in each phase probability curve to be the phase, one would obtain the most probable phases, and therefore the most probable electron density. However, to get the best estimate of phases, it is necessary to find the phases with least square error. In order to minimize this error, the center of gravity of the curve is chosen as the phase, and to further improve maps the resulting structure factor is weighted by the likelihood of the phase being correct. That weighting factor is the figure of merit, m, which is a number between 0 and 1 defined by:

$$m = (\ P(\alpha) \exp(i\alpha)\, d\alpha) / (\int P(\alpha)\, d\alpha) = (\sum P(\alpha) \cos(\square\alpha)) / (\sum P(\alpha)) = \langle \cos \square\alpha \rangle$$

This value, m, is the mean value of the cosine of the error in the phase angle for a given reflection. The figure of merit is used in finding the best value of $\mathbf{F}_{hkl}$ to use in phasing by:

$$\mathbf{F}_{hkl}(\text{best}) = m\,|F_{hkl}|\,\exp[i\alpha(\text{best})]$$

The electron density maps calculated from $\mathbf{F}_{hkl}$(best) can be further improved by density modification (DM) techniques (Brunger *et al.* 1998). The types of DM used in the solution of MtaN were histogram matching (Lunin 1988) and solvent flipping (Abrahams 1997). Solvent flipping is a minor variation of solvent flattening. Because proteins have a high electron density compared to solvent, the initial electron density map will have large areas of high density and other areas of low density. By estimating the solvent volume of the unit cell, it is possible to artificially set the density of the lower density areas (which should be solvent) to a low constant value and recalculate the structure factors based on the solvent flattened map. The new phases and observed amplitudes are used to recalculate the map, which now has more clearly defined high-density areas. Solvent flipping does essentially the same, except that instead of flattening solvent features to a low constant value, the signs of the low-density values are reversed. This has the advantage of reducing bias when recombining the data from the observed amplitudes and solvent flattened (flipped) phases, and is therfore a more powerful technique. Both solvent flipping and solvent flattening work better when a protein crystal has a high level of solvent in the unit cell. Histogram matching was also used on MtaN maps (Lunin 1988). This is possible because correct protein electron density will have a

predictable distribution of density values. This does not require modifying areas of the electron density map, but rather the distribution of density values calculated. A histogram of protein electron density is essentially independent of the protein structure or sequence, and one will look like another. In histogram matching, the histogram of density grid points in the calculated map is matched to the known regular protein density histogram. This is done by applying a scaling factor and a shift factor to each data bin. The adjusted density points are then used to recalculate the structure factors, and the new phases are used in the calculation of new electron density maps. Density modification is a powerful tool in the improvement of phases and production of interpretable maps.

Model Building, Refinement and Validation

Once maps have been calculated, a protein model is built to match the maps, and therefore the observed intensities data. In addition to the electron density maps, the crystallographer has knowledge of the basic structures of proteins and the amino acid sequence of the protein to be built into the model. In the case of MAD phasing (of MtaN), locations of selenium atoms are also known, from which selenomethionine amino acyl residues can be built. These residue locations are used to begin to put the protein chain in the correct register. Early in the fitting process, the maps are based only on the MAD phasing data. They can still be noisy and some parts of the molecule(s) appear to be missing because the calculated phase angles are still poor. The first step in model building is to locate and build the backbone of large sections of secondary structure, since these areas are usually well-defined in the maps and the general backbone configuration of these elements is predictable. Once obvious sections of backbone have been traced,

attempts are made to connect the backbone pieces to each other through the less predictable turns and loops. When a backbone has been traced, the process of fitting side chains begins. Some side chains will be obvious, such as large, buried aromatic residues. Selenomethionine residues will be particularly obvious since their selenium atoms contain many more electrons than the average protein atom and their locations have already been determined in the phasing step. These obvious residues are helpful in figuring out which side chains belong in the less obvious positions. At first, the model is built only to match the observed density in the initial electron density map. Phase combination, a process of adding new phase information from the growing model to the calculated phases, can also be used to improve the electron density map once a model has begun to take shape (Bricogne 1976).

Once an initial model is built, it is refined against the observed intensities data. In practice, a model will never perfectly match the data, but a measure is available of how closely it matches. The R-factor is defined by:

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

and is expressed as a percentage. F_{calc} is the calculated structure factor amplitude ($|F|$) if the model were 100% correct. It provides a measure of how well (or poorly) a model matches a data set by comparing the difference between the calculated structure factors (from the model) and the observed structure factors. It is not uncommon for this number

to be as high as 50% for an initial model, even though 59% would be expected from a random structure. The final working R-factor (R_{work}, described later) of a good quality model should be below 25%.

Refinement is the process of bringing the model to more closely match the data and known protein properties. At resolutions greater than ~1.0 Å, this is done by the use of stereochemical restraints. The bond lengths and angles for all the bonds which will be found in a protein are known quite precisely from small molecule studies. They have been confirmed through the years by high-resolution protein structures, including the 0.54 Å resolution structure of crambin, a plant protein from *Crambe abyssinica* that is homologous to plant toxins, which is the highest resolution protein structure in the protein data bank (Jelsch *et al.* 2000). Refinement brings the atom positions into the closest possible agreement with both of these considerations while remaining in agreement with the observed intensities data.

Bringing the model into agreement with the observed data is done generally through the optimization of the function:

$$Q = E_{restraints} + \Sigma_{hkl}\, w_{(hkl)}\, \{|F_{obs(hkl)}| - |F_{calc(hkl)}|\}^2$$

where $w_{(hkl)}$ is included as a weighting factor, and $E_{restraints}$ is a term relating the geometry of the model to known geometry such as bond lengths, bond angles, planarity of bonds, and non-bonded contacts. Adding the $E_{restraints}$ term keeps the model from deviating too far from ideal geometry. The atoms are "moved" around their positions slightly, and the

above function is recalculated. By adjusting $w_{(hkl)}$ and the weight of the geometric restraints, the refinement can be tuned to give more weight to geometry or observed structure factor amplitudes. After some cycles of refinement have converged, a new electron density map is calculated with newly calculated phases and the model is adjusted to fit the new map by the crystallographer. As the model improves, the model phases improve and early in refinement, the phases calculated from the model will be more correct than the phases from the "heavy atom" data. As the cycles of refinement and rebuilding continue, the model will continue to improve and grow and more closely match the observed data, thus the R-factor will drop.

The computer program suite CNS was used in the structure determination of MtaN (Brunger *et al.* 1998). It contains a variety of programs that can be used in many of the common tasks of macromolecular crystallography. Refinement in CNS works by combining simulated annealing (Brünger *et al.* 1990), energy minimization (Pannu *et al.* 1996), and B-factor refinement (described below). Simulated annealing is a process where the input model is "heated" and slowly "cooled", which means that the bonds are allowed to have a higher energy (temperature) at first while atom positions are refined against observed data, then the model is "cooled" by allowing the bond lengths and angles to have slightly less energy, again while the atom positions are refined against the observed data. This process repeats until the bonds are in low-energy configurations. Energy minimization brings non-covalent interactions, such as hydrogen-bonds, to low-energy states, again while refining against observed data.

These cycles of refinement and rebuilding generally begin with the lower resolution data because they are generally measured more accurately and completely. They also allow greater initial freedom in refinement of a poor initial model because the radius of convergence (positional range that atoms are allowed to sample during refinement) is a function of the resolution of the data used. Using data of lower resolution than ~3 Å requires that only the positional parameters (x, y, z) of each atom be refined. However, later in the process, as the model improves and higher-resolution data are included, another parameter, the thermal factor (B), is included in the refinement which accounts for the motion of each atom to some degree. As an X-ray passes through a crystal, identical atoms will not be at exactly the same positions within different unit cells. This is, in part, because those atoms are moving slightly about their equilibrium positions, in effect creating scattering from a "smeared" atom. The measure of the motion of an atom is its B-factor and is given by:

$$B = 8\pi^2 <\mu^2>$$

Where $<\mu^2>$ is the mean square displacement of the atomic vibration, thus B factors are expressed in units of $Å^2$. Atoms with higher B factors diffract less strongly because of their increased thermal motion, so the B factor is used in calculating the atomic scattering factor (f) of each atom.

As the model and phases improve, portions of the molecule which were not visible in early maps will appear and be modeled into the new density. Ordered water molecules will also appear in the density maps, and adding them into the model further

improves the model and the R-factor. In general, the higher resolution the map becomes, the more ordered water molecules will become visible and the better the positioning of the atoms of the protein. Iterations of model building and refinement will eventually lead to insignificant improvement in the model, at which time the model is "done".

Assessing the quality of the final model is done in many ways. The first, mentioned above, is the R-factor. A variation of this is called the R-free. Because the same observations used in refinement of the model are also used in calculation of the R-factor, they represent a biased measure. The R-free is calculated as the R-factor, except that it is done using observations which were not included in the refinement process. Normally 5% to 10% of the observed data are set aside for this purpose and the rest of the data are used in a "working set", from which R_{work} is calculated. This makes the R-free an unbiased R-factor (Brünger 1992). Another check of the model can be made through "omit maps". These maps are made by deleting a part of the model (5% to 10%), refining it slightly, then calculating structure factors based on the partial model, which are then used to recalculate the maps. The recalculated maps were made without the bias of the deleted section, so if the density of that section appears in the omit maps, one can be more assured of the validity of that portion of the model. Omit maps are done through the entire model, a process which can now be done automatically in CNS to produce "composite omit maps."

Perhaps the most common check on the stereochemistry of a model is the Ramachandran plot (Figure 2.20). The Ramachandran plot is a graphical representation

of the ϕ - ψ angle combinations found in a model, and includes areas on the plot which are commonly found and those which are disallowed. Glycine residues, because of their unique structure (lacking a β carbon), may fall in areas of the plot larger than the other amino acids. The main types of secondary structure fall into specific areas of the plot. A correct model should have the majority of the residues in the most favored regions and none (except glycines) in the disallowed regions. Sometimes a correct model will have a few residues in these areas because those residues are in unusual areas of the protein, perhaps a ligand binding site or enzyme active site.

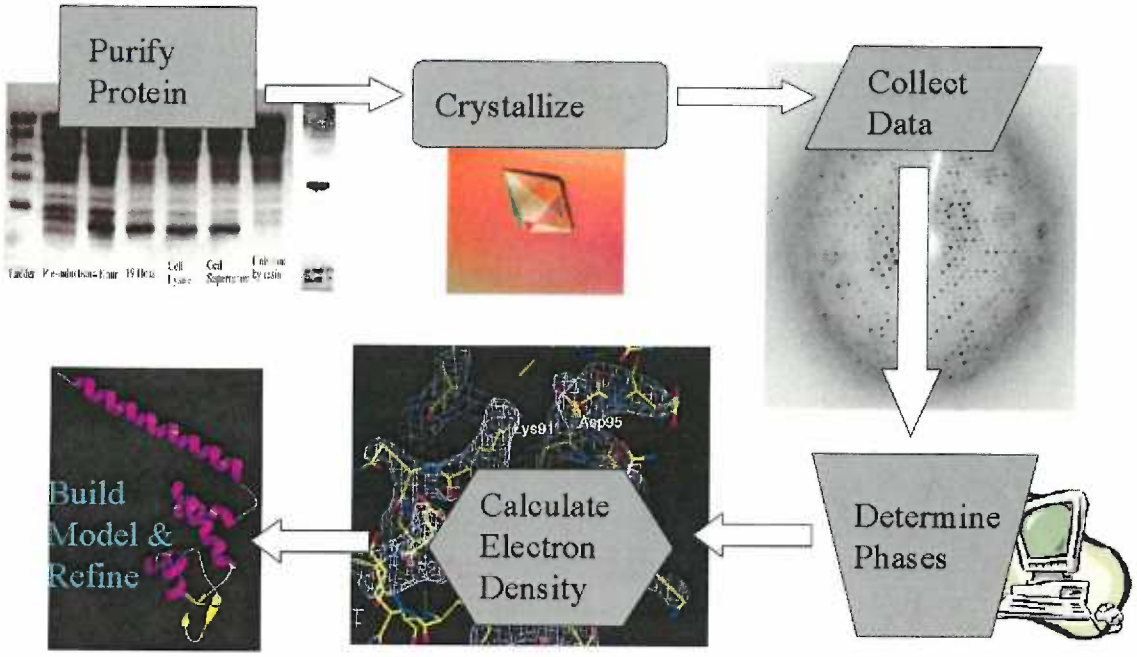

Figure 2.1 A graphical flow-chart of the process of macromolecular structure determination. All images used are from the *de novo* structure determination of MtaN.

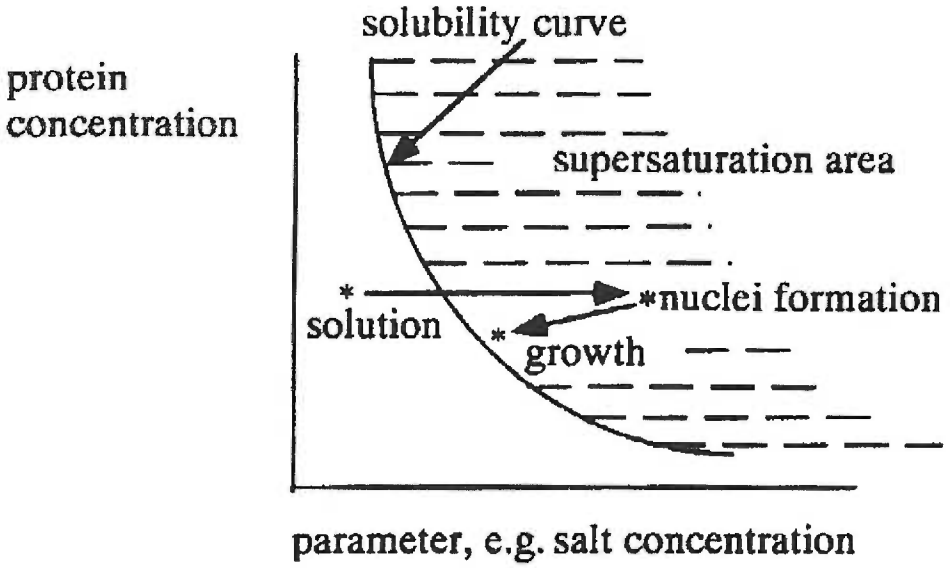

Figure 2.2 A solubility diagram illustrating the stages in protein crystal formation. In an ideal experiment, a protein will begin in the soluble area of the diagram and as the precipitant concentration increases the protein will shift into the supersaturation area. At this time a nucleus will form and crystal growth will begin. As crystal growth proceeds, the concentration of protein will decrease slightly as molecules come out of solution in the growing crystal, keeping the protein concentration on the edge of the solubility curve. From (Drenth 1999).

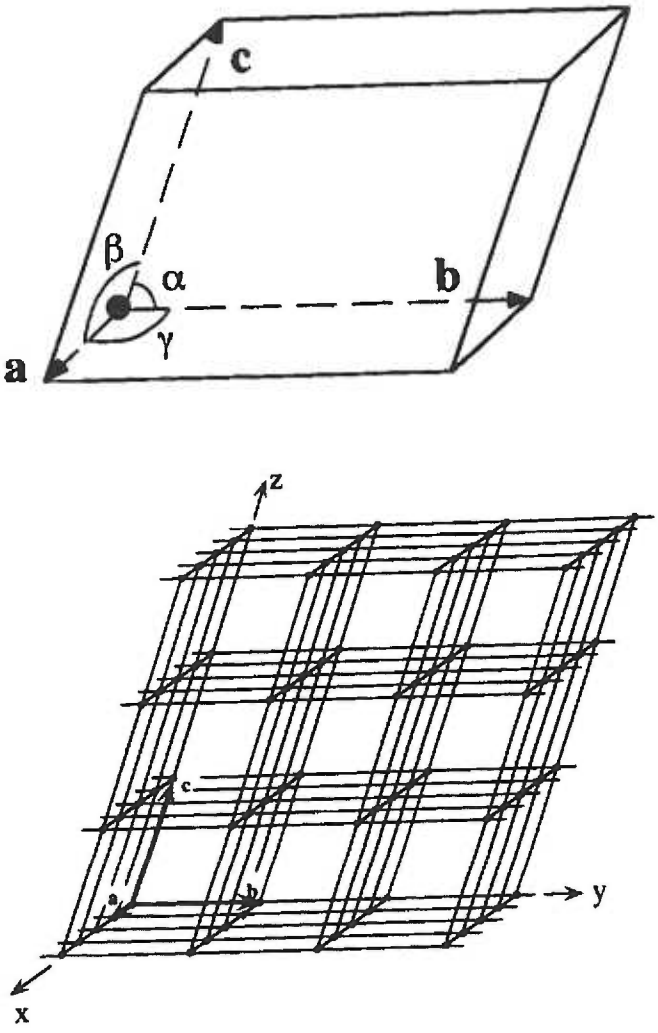

Figure 2.3 Top) A unit cell. Edges of the unit cell are defined by **a**, **b** and **c**, and the angles between them are α, β, and γ. Bottom) A lattice built of many unit cells. The lattice is merely a replication of translated unit cells. From (Drenth 1999).

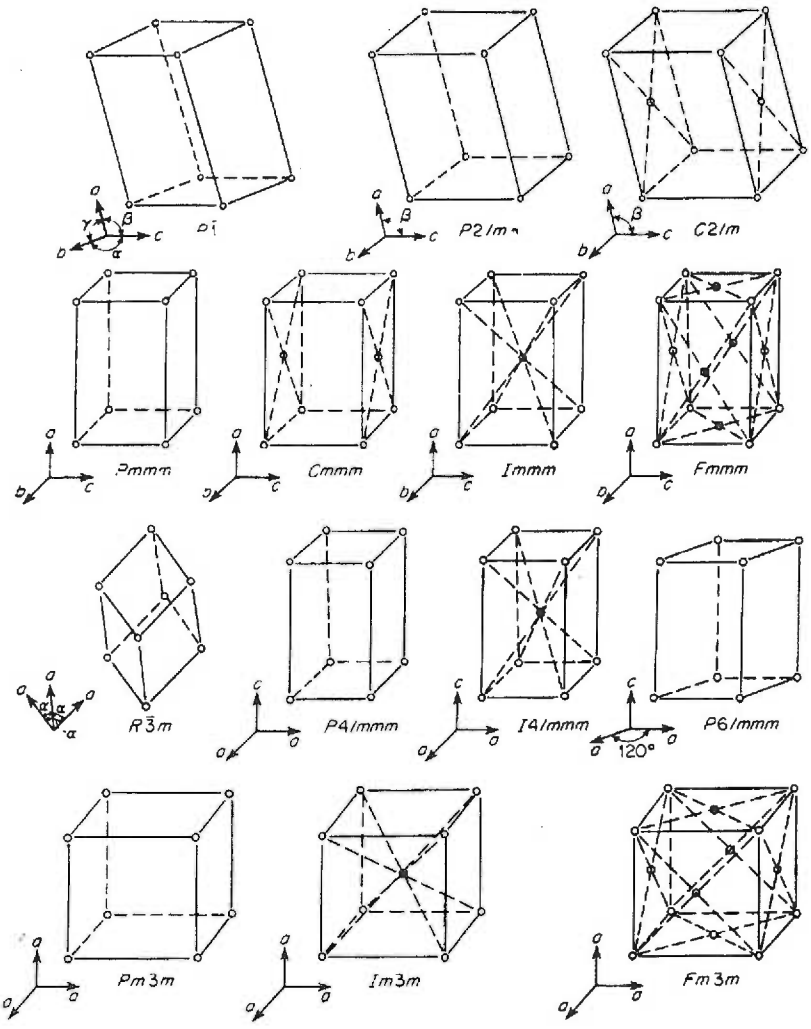

Figure 2.4 The 14 Bravais lattice types, including examples of primitive cells (P) with lattice points only at the corners, C face-centered cells (C) with lattice points at the corners and on the C face, all-faces centered (F) with lattice points at the corners and all faces, and inner-centered (I) with lattice points at the corners and in the center of the cell. From (Stout *et al.* 1989).

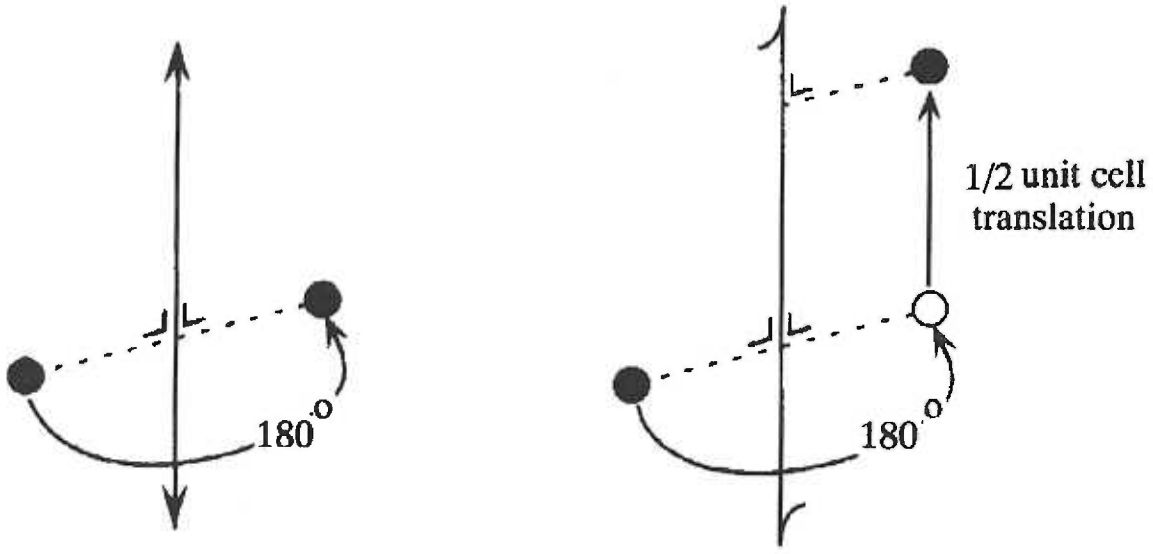

Figure 2.5 A two-fold axis (left) is a simple 180° rotation around the axis. A two-fold screw axis (right) includes a translation of ½ unit cell along with the rotation. These screw axes are denoted by the 2_1 in $I2_12_12_1$, the space group of MtaN. From (Drenth 1999).

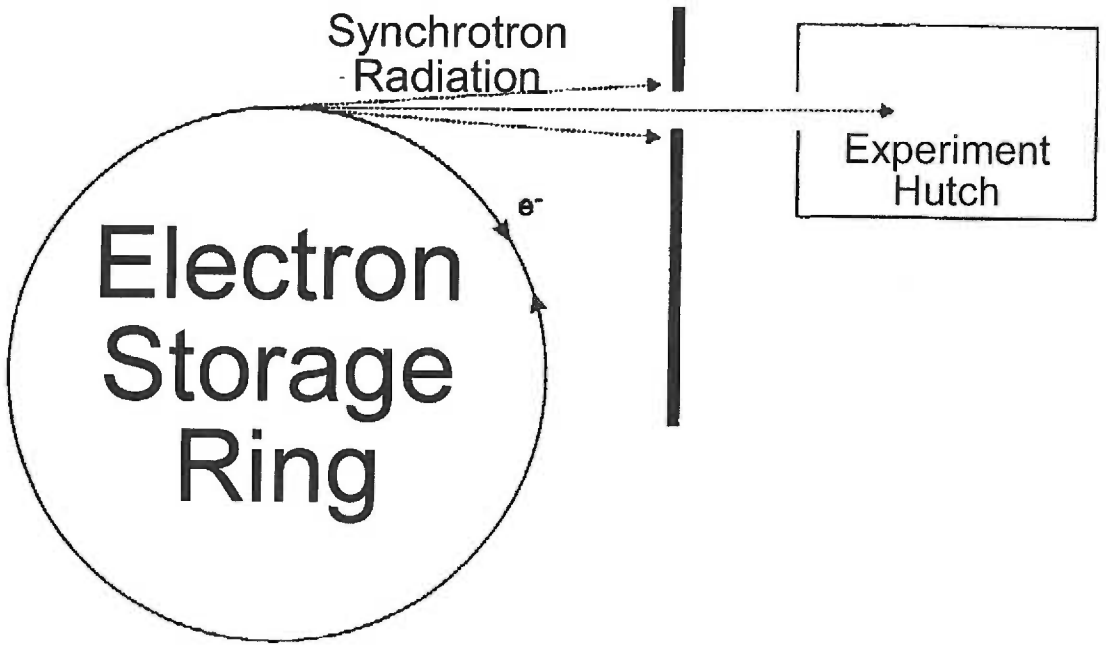

Figure 2.6 A schematic of a synchrotron radiation source. Electrons are moving around the storage ring in a circle. As they accelerate around the ring, they emit radiation. Experiments are conducted in the hutch, which is well-shielded to protect researchers. Adapted from (McRee 1999).

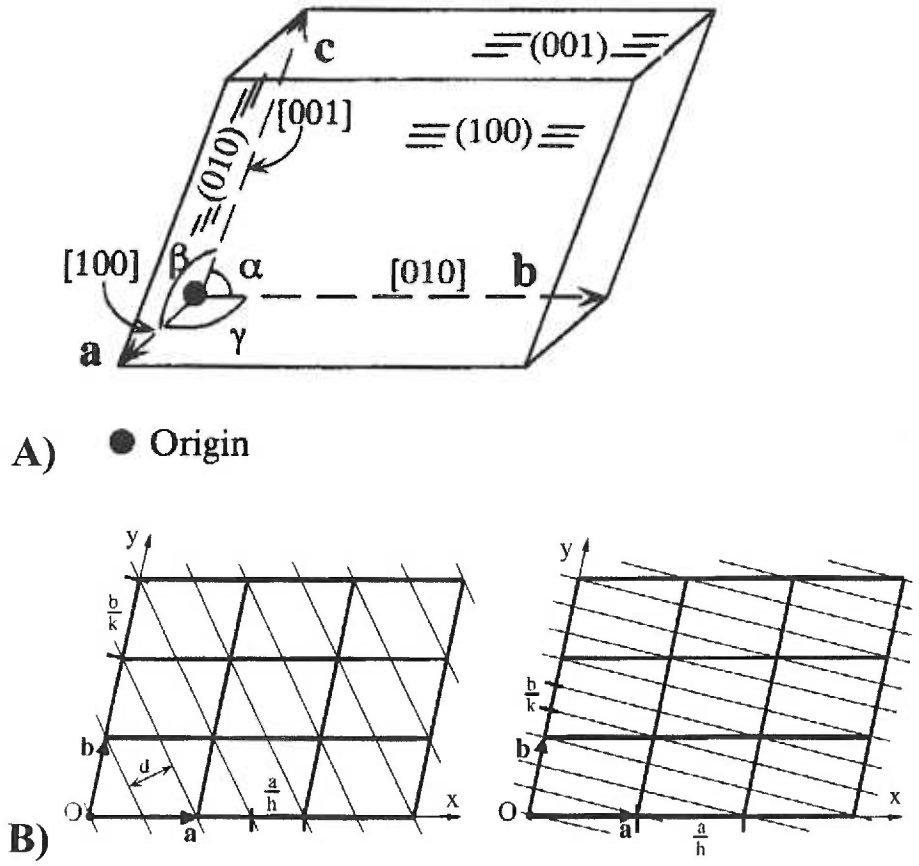

Figure 2.7 A) A unit cell with the edges (in brackets) and planes (in parentheses) labelled with their respective *hkl* indices. B) A two-dimensional picture of a lattice in the **ab** plane. In the figure on the left, the 21 *hk* lines are shown, because they cut **a** in half and they only cross **b** at the lattice points. The right figure shows the 1,3 *hk* lines; they cross **a** only at lattice points (cut **a** only once) but cut **b** 3 times. The d in the left figure is the d-spacing (perpendicular distance) between those lines. The *hkl* planes are constructed the same way in three dimensions, and the d-spacing is the distance between adjacent planes. From (Drenth 1999).

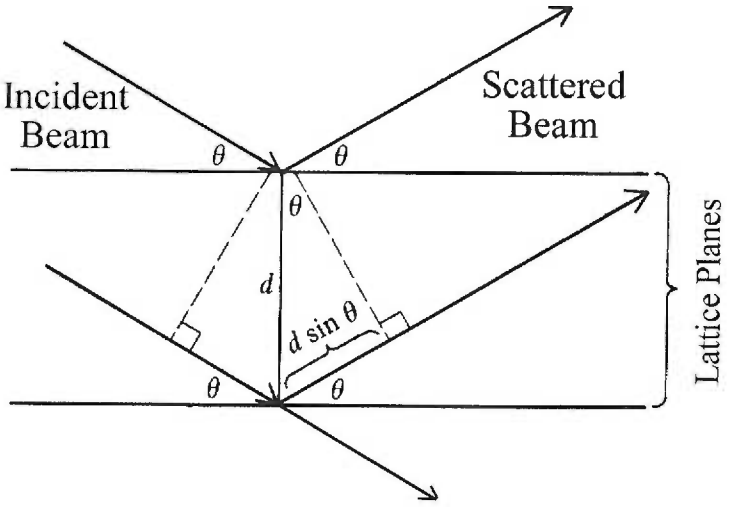

Figure 2.8 Bragg's law ($2d \sin \theta = n\lambda$) requires that the extra distance travelled by a wave be a multiple of the wavelength. This diagram shows that relationship is satisfied only when the extra distance travelled by the second incident beam = $n\lambda$. Because diffraction is measured with n =1, and λ will remain constant throughout an experiment, $\sin \theta$ is proportional to $1/d$. By measuring 2θ, the angle between the incident and diffracted beams, the d spacing, which is the perpendicular spacing between planes of reflection, can be determined. From (Cantor *et al.* 1980).

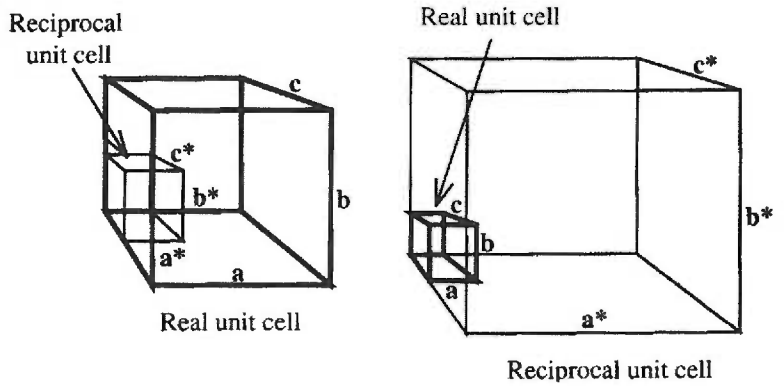

Figure 2.9 Real and reciprocal unit cells. These are orthorhombic cells, so that the reciprocal edges are parallel to the real edges, but this is not the case when angles other than 90° are present in the regular unit cell. A smaller real cell leads to a larger reciprocal cell and vice-versa. From (Rhodes 1993).

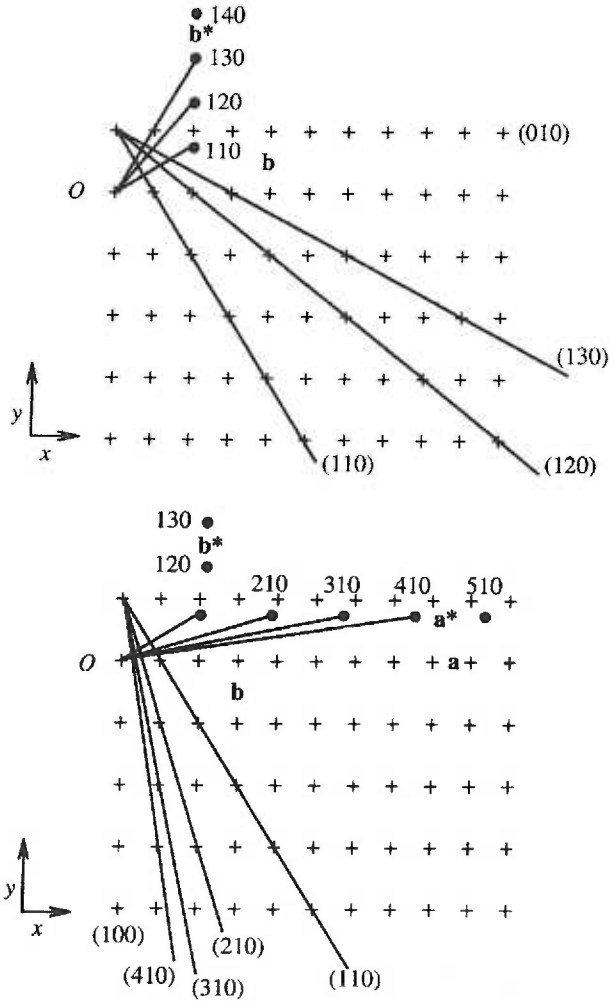

Figure 2.10 Construction of the reciprocal lattice. The real lattice points are shown as plus signs and the reciprocal lattice points are dots. An origin is chosen, then a line is drawn perpendicular to a reflecting plane with one end on the origin and of length 1/d, the inverse of the perpendicular d spacing for that set of planes. The end of that line is given the designation of the *hkl* planes from which it originated. The case where l = 0 is shown above, however, the process continues until a 3-dimensional lattice is completed. From (Rhodes 1993).

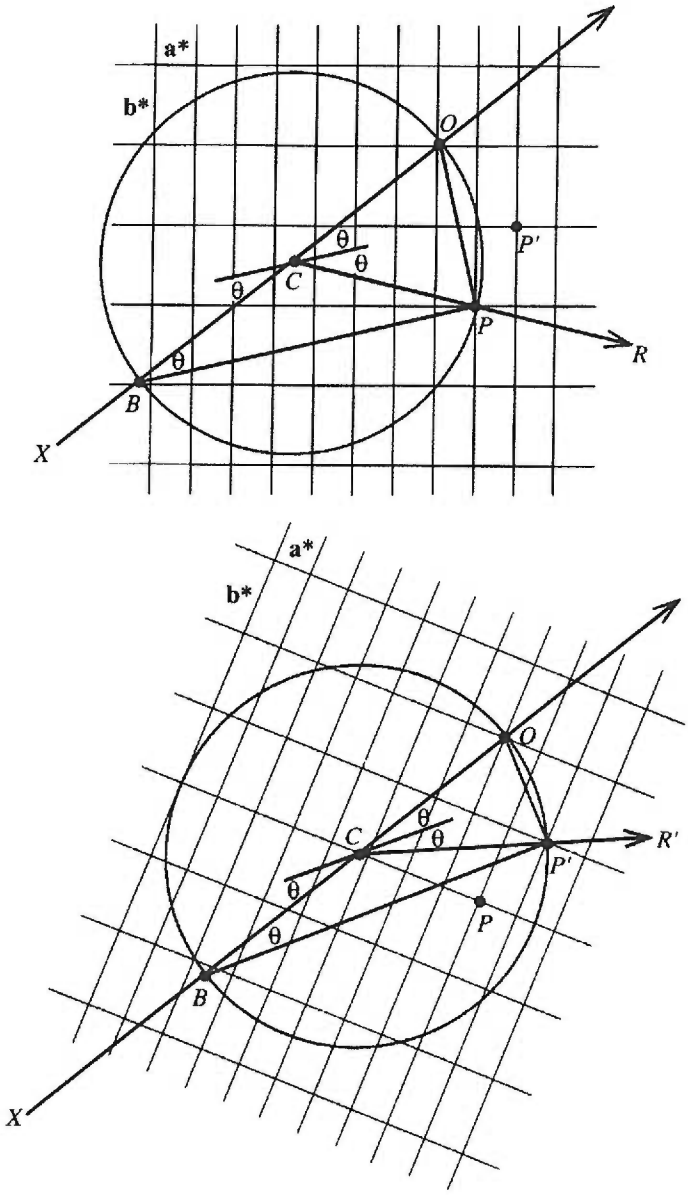

Figure 2.11 Diffraction in reciprocal space. Diffracted ray *R* (top) emerges from the crystal at angle 2θ when reciprocal lattice point *P* touches the *circle (sphere) of reflection*. The reciprocal lattice rotates with the crystal until point *P'* touches the circle producing ray *R'* (bottom). From (Rhodes 1993).

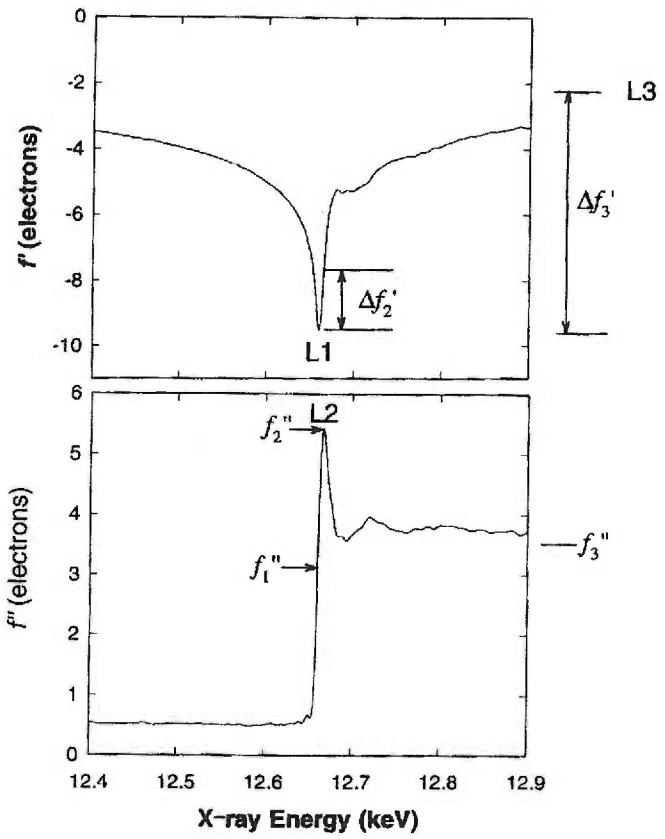

Figure 2.12 Variation of the real (f , top) and imaginary (f , bottom) components of the anomalous scattering around the K edge of selenium. Wavelengths are chosen at the bottom of the f curve (L1), the peak of the f curve (L2), and at a remote wavelength (L3, to optimize f and f effects). The difference signals ($\Delta f_2 = f_2 - f_1$ and $\Delta f_3 = f_3 - f_1$) are also maximized by this method . Adapted from (Ramakrishnan *et al.* 1997).

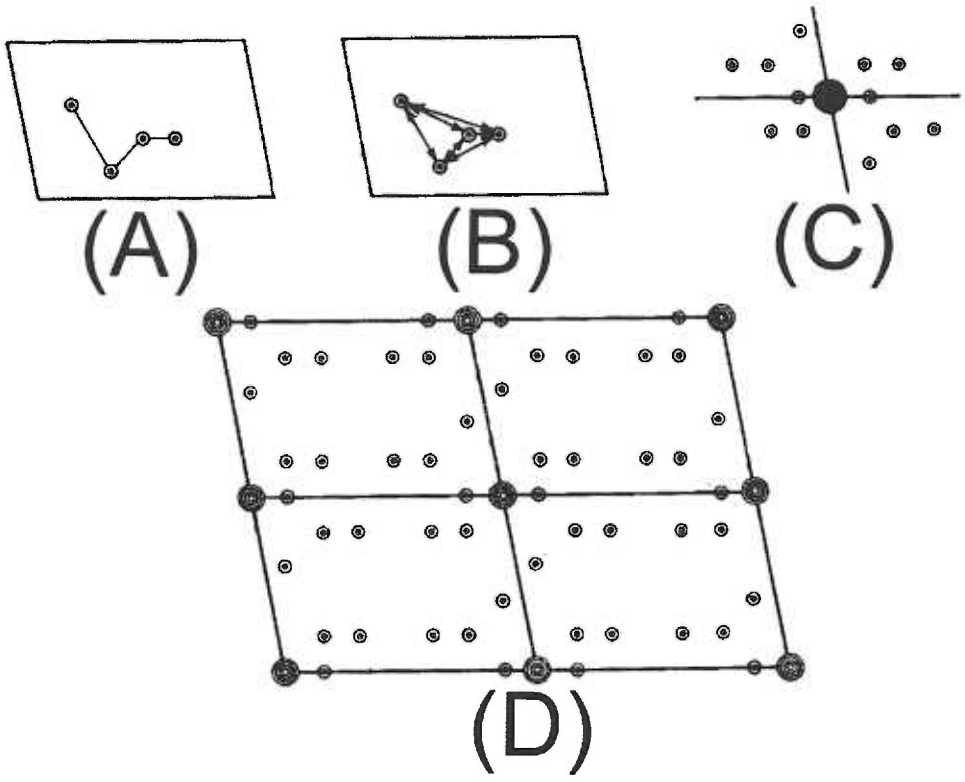

Figure 2.13 Construction of a two-dimensional Patterson map. (A) Four "atoms" in a structure. (B) The atom-to-atom vectors which will become the Patterson "atoms". (C) The tails of the vectors from B are placed on the origin to produce the 12 Patterson peaks. (D) Four Patterson cells constructed by translating the peaks a unit cell in each direction. As described in the text, heavy atom Patterson peaks will be much higher than those of the lighter atoms, enabling a crystallographer to locate them in the Patterson unit cell. From those locations, their real unit cell locations can be determined. From (Stout *et al.* 1989).

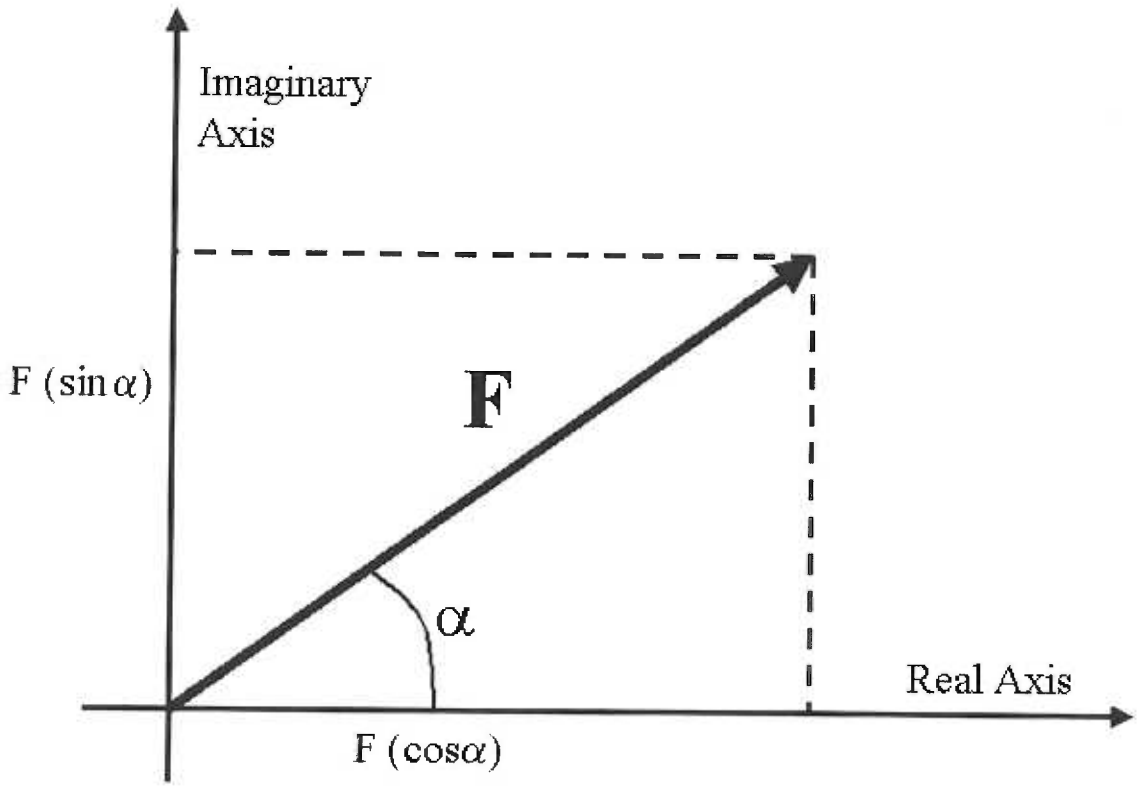

Figure 2.14 An Argand diagram showing the phase (α), real component (F (cos α)), and the imaginary component (F (sin α)) of the vector **F**. Adapted from (Drenth 1999).

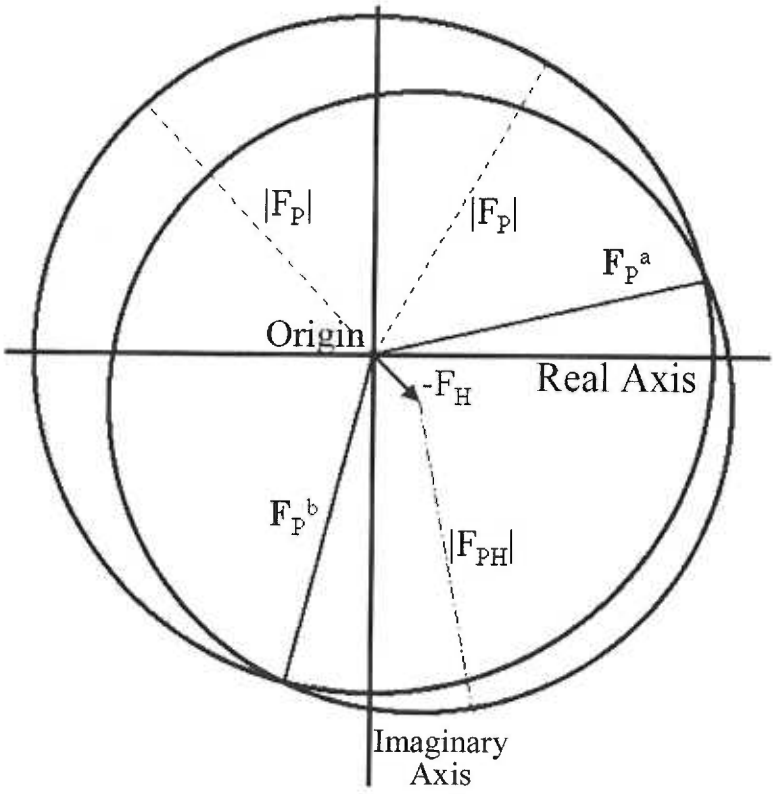

Figure 2.15 Phase determination by the Harker construction. $\mathbf{F}_P^b$ and $\mathbf{F}_P^a$ are the two possible structure factors from this construction. Described in the text.

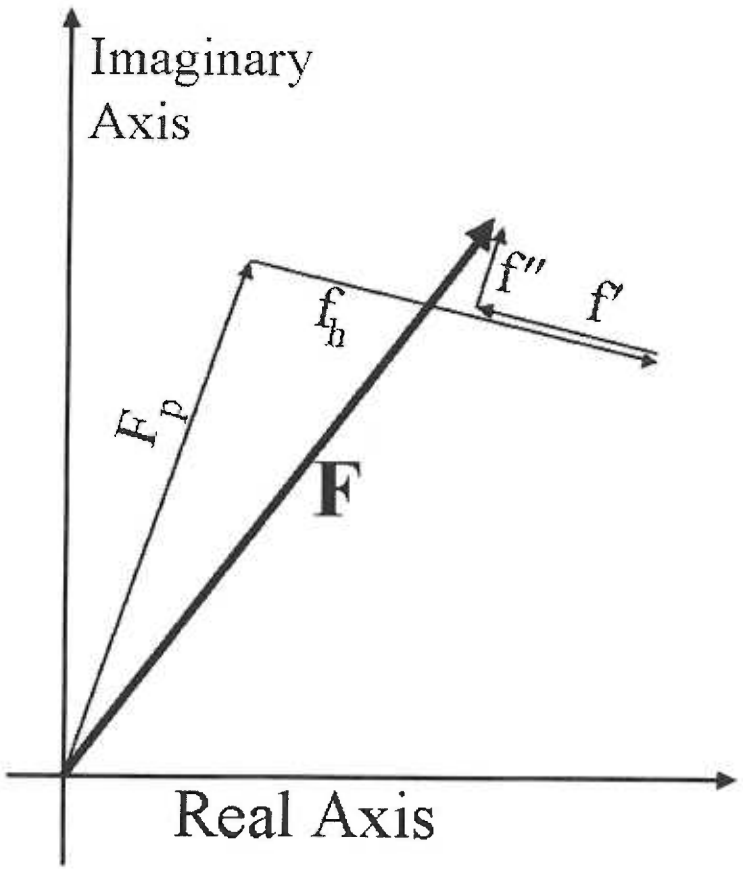

Figure 2.16 The additive effects of scattering from the non-anomalous protein atoms (F_P), the regular scattering from the "heavy atoms" (f_h), and the real (f) and imaginary (f) components of the anomalous scattering from the "heavy atoms". Adapted from (Phillips *et al.* 1980).

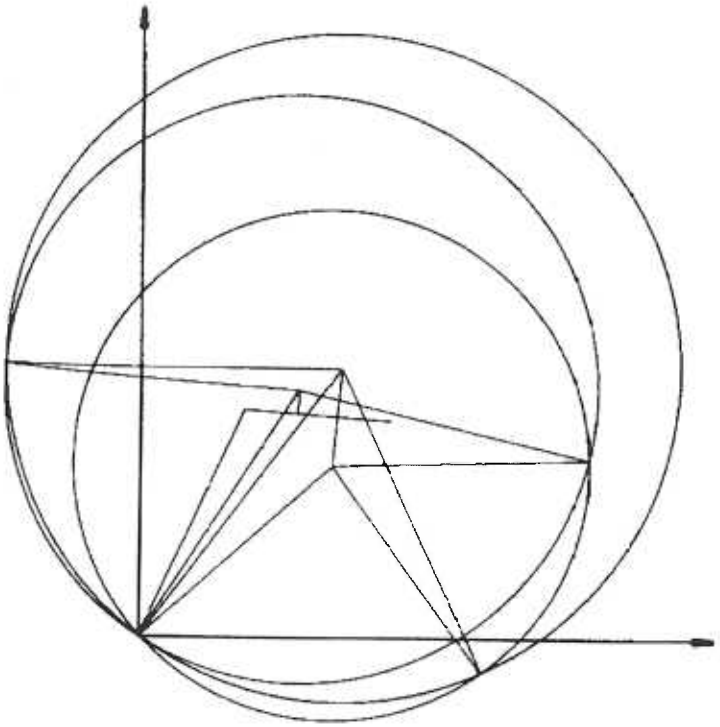

Figure 2.17 A Harker construction using two wavelengths and a Bijvoet mate from one of the wavelengths. From (Phillips *et al.* 1980).

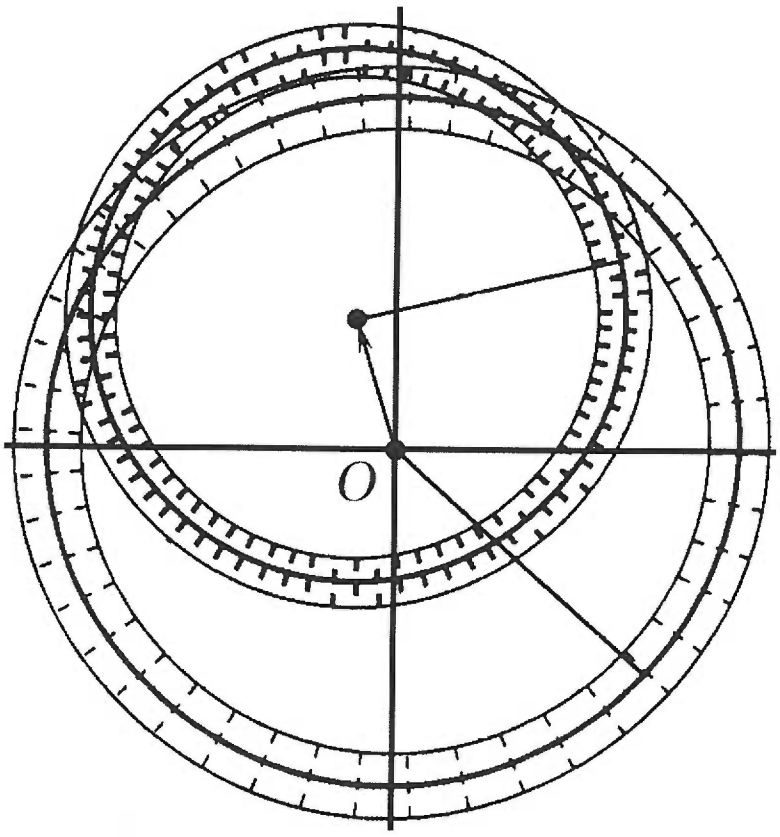

Figure 2.18 "Density" in circles of the Harker construction due to uncertainty in measurement of intensities. The regions of overlap are multiplied to obtain a probability function for the correct phase. Adapted from (Drenth 1999).

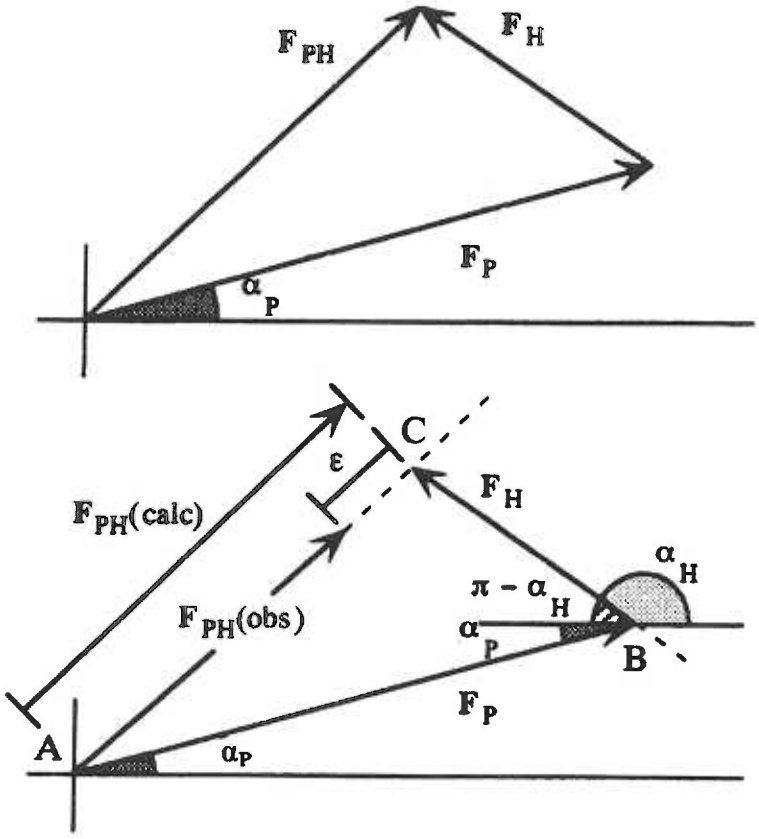

Figure 2.19 The Lack-of-closure method. The top situation is the ideal case, where the triangle closes perfectly for $\mathbf{F}_P$ (observed), $\mathbf{F}_{PH}$ (observed) and $\mathbf{F}_H$ (calculated). The lower illustration shows a more realistic situation, where the triangle does not close and the observed and calculated values of $|F_{PH}|$ differ by the lack of closure error, ε. From (Drenth 1999).

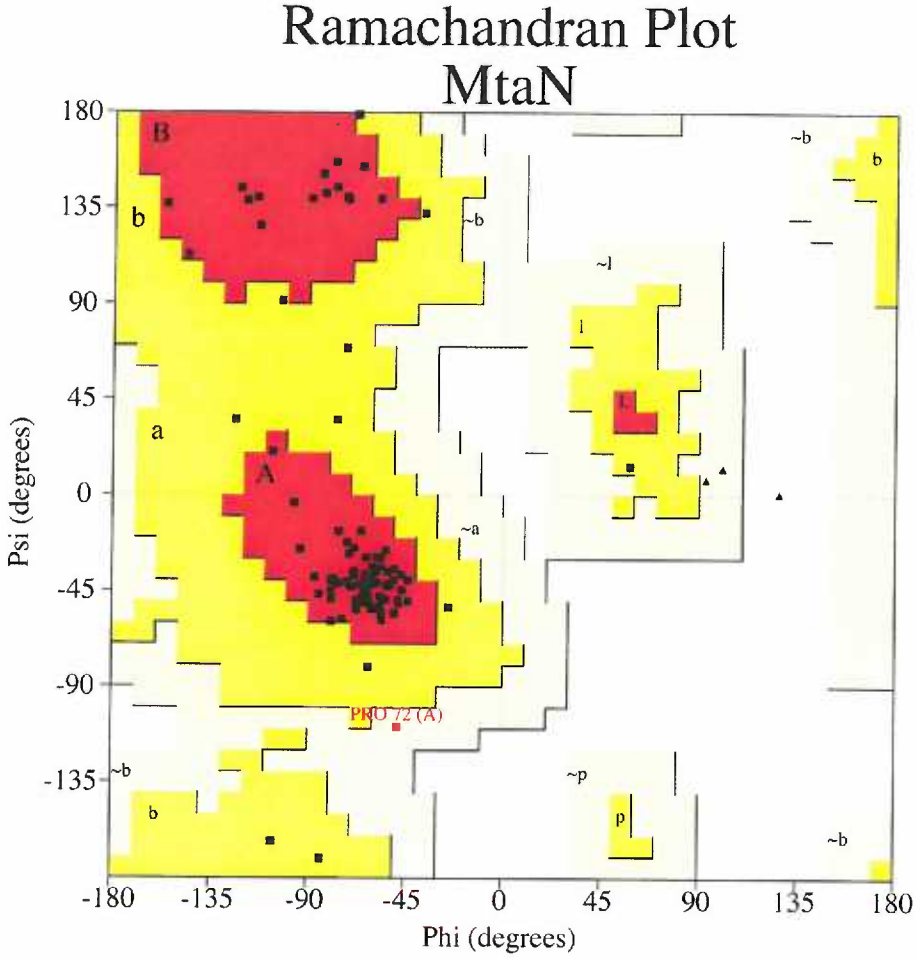

Figure 2.20 A Ramachandran plot of the final MtaN model. All non-glycine residues are in the most favored or allowed regions except for Pro72, which is in the generously allowed area. The majority of the residues cluster in the α-helix portion of the plot, labelled A and a. The β-sheet portions of the plot are labelled B and b, and the left-handed helix portion of the plot is labelled L and l. Red areas are most favored, and white areas are disallowed for all residues except glycine, with intermediate areas colored yellow. Created by PROCHECK (Laskowski *et al.* 1993).

Chapter 3

CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION STUDIES ON THE DNA-BINDING DOMAIN OF THE MULTIDRUG TRANSPORTER ACTIVATION PROTEIN (MTAN) FROM *BACILLUS SUBTILIS*

Michael H. Godsey[a], Natalya N. Baranova[b], Alexander A. Neyfakh[b] and Richard G. Brennan[a*]

[a]Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon, 97201-3098, U.S.A

[b]Center for Pharmaceutical Biotechnology, University of Illinois, Chicago, Illinois, 60607, U.S.A.

*To Whom Correspondence Should Be Addressed

Acta Cryst. (2000). **D56**, 1456-1458.

Abstract

The N-terminal DNA-binding domain of the multidrug activation protein (MtaN) was crystallized by the hanging drop vapour diffusion method using lithium chloride as a precipitant. The crystals are orthorhombic and belong to the space group $I2_12_12_1$ with unit cell parameters a = 49.4, b = 67.8, c = 115.0 Å. Diffraction data have been collected at 100K to 2.75 Å resolution at a synchrotron radiation source.

Introduction

Bacteria express a variety of membrane-bound proteins that extrude toxins from the cell. Because the recognized compounds include many antibiotics, these proteins have been named multidrug transporters (Paulsen *et al.* 1996; van Veen *et al.* 1999). In *Bacillus subtilis*, two of these proteins, Bmr and Blt, have been shown to be controlled by specific transcriptional regulators, BmrR and BltR, respectively (Ahmed *et al.* 1994; Ahmed *et al.* 1995). However, both transporters are further regulated by the multidrug transporter activation protein (Mta) (Baranova *et al.* 1999). This protein also controls its own transcription and that of at least one more gene, *ydfK*, which encodes a putative membrane protein.

Mta, like BmrR and BltR, is a member of the MerR family, named for the mercury-resistance gene regulator (Summers 1992; Hidalgo *et al.* 1997b). This family has high sequence conservation in the N-terminal, DNA-binding domain, which is predicted to contain a helix-turn-helix DNA-binding motif. However, the C-terminal domains of the family are widely divergent in both size and sequence. Beyond the

helix-turn-helix motif, there appears to be no sequence or structural homology between the members of the MerR family and any other gene regulators.

Mta contains 257 residues and is a functional dimer. A truncated form of Mta, encompassing the first 109 amino acids (MtaN), is able to bind to the promoter elements and activate the transcription of genes under control of the full-length protein (Baranova *et al.* 1999). Thus it appears the C-terminus contains a regulatory domain, removal of which is sufficient to relieve inhibition of transcriptional activation. *In vivo*, MtaN activates its own transcription, creating a positive feedback loop leading to high levels of intracellular protein. This may be important, as our binding studies have found that MtaN has 30- to 50- fold higher K_d values for the *bmr* and *blt* promoters respectively, than for the *mta* promoter (data not shown).

Our interest in MtaN is twofold. The promoters targeted by MerR family members contain an unusually long 19 base-pair spacer between the -10 and -35 elements, and the structural mechanism of transcriptional activation by this family remains unclear. In addition, the known promoter targets of MtaN are quite dissimilar, with only 12 consensus base pairs in a stretch of 31, of which only six are found in the DNase I footprint of the protein (Baranova *et al.* 1999). Such limited base-pair conservation suggests an unusual mechanism of multiple promoter recognition.

To understand the molecular mechanism of transcription activation by MerR family members and multiple promoter recognition by a single protein, we have undertaken experiments to determine the three-dimensional structure of MtaN by X-ray diffraction methods. In this paper, we report the crystallization and preliminary X-ray diffraction analysis of the DNA-binding domain of the MerR family member Mta.

Experimental Results and Discussion

Purification and solubilization.

Expression and purification of MtaN in *Escherichia coli* was performed essentially as described in Baranova *et al.*, (1999). One litre of cells expressing MtaN fused to a C-terminal six-histidine tag was pelleted, resuspended in ~25 ml of lysis buffer (20 m*M* Bis-Tris-Propane pH 8.9, 250 m*M* NaCl) and lysed in a French pressure cell at 10 000 psi. Lysates clarified by centrifugation were added to 4 ml of a 50% slurry of Talon cobalt-affinity resin (Clonetech) and purified by a hybrid batch-gravity flow method. Briefly, the clarified lysates were mixed with resin slurry, allowed to bind and washed several times. The resin was then poured into a small gravity-flow column for stepwise elution. Protein was eluted with 50 m*M* imidazole and fractions containing MtaN were combined and dialyzed against storage buffer (20 m*M* Bis-Tris-Propane pH 8.9, 100 m*M* NaCl). The solubility of the protein was found to be inadequate for crystallization below pH 8.5. Subsequent dynamic light-scattering experiments indicated enhanced solubility and monodispersity at pH values ≥ 8.9. The protein was concentrated to ~11 mg ml^{-1} in a Centricon 3 and used for crystallization experiments. One litre of culture yields approximately 20 mg of pure protein.

Crystallization

Crystallization conditions were screened by vapour diffusion using the hanging-drop method. Droplets which contained 2 μl each of protein and reservoir solution were equilibrated over 1 ml of reservoir solution at room temperature. Crystals were obtained in 2 days from a reservoir solution of 3.77 *M* LiCl, 50 m*M* HEPES pH 7.0, 10

m*M* $MgCl_2$. A 20 µl drop yields one or two crystals with dimensions of approximately 0.4 x 0.4 x 0.6 mm in a few days (Figure 3.1). These crystals were found to be too unstable for data collection at room temperature. In one attempt to lengthen the lifetime of these crystals, drops containing single crystals were removed from their wells, placed over reservoirs containing 5.0 *M* LiCl in the crystallization buffer and allowed to re-equilibrate. Surprisingly, these crystals typically doubled in size overnight, and their lifetime in the X-ray beam was increased enough to allow the determination of unit cell parameters and limit space-group choice to one of two. However, for complete data collection, the crystals must be frozen. A systematic search for a good cryo-protectant revealed that the crystallization solution alone, containing 3.77 *M* LiCl, is quite effective for protection.

X-ray diffraction

Preliminary intensity data for unit-cell parameter and space-group determination were collected on an R-Axis IV image-plate system equipped with double-focusing mirrors and using Cu K_α radiation from a Rigaku RU300 rotating-anode generator operated at 50 kV and 100 mA. These data revealed an orthorhombic space group, either *I*222 or $I2_12_12_1$, which are indistinguishable from each other based on diffraction patterns. A V_m value of 3.6 Å Da^{-1} (Matthews 1968) and a solvent content of 63% are consistent with single molecule in the asymmetric unit. There is a band of diffuse scattering in the diffraction pattern at resolutions between 4.0 Å and 3.0 Å (Figure 3.2). The cause of this is unclear, although it is likely to be a consequence of thermal motion in the crystal lattice. All experimental attempts to eliminate this feature have failed.

However, the thermal diffusion does not seem to interfere significantly with successful data collection and intensity integration.

Although the crystals are very unstable in the X-ray beam at room temperature, they are extremely stable in the crystallization drop. As an example, the best intensity data have been collected from a crystal over 18 months old (Table 3.1). For intensity data collection, a crystal was removed directly from the drop in which it was grown, flash-frozen in the N_2 stream of an X-stream cryo-cooling system (Molecular Structure Corporation) and stored until data collection. Intensity data were collected at 100 K on beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) using a MAR 345 detector and $\lambda = 1.08$ Å. Intensity data were processed and scaled using *MOSFLM* (Collaborative Computational Project 1994).

Heavy-atom soaks have been unsuccessful in derivatizing the protein, although this may be in part be a consequence of the presence of the high salt concentration in the crystallization conditions. As an alternative strategy, selenomethionyl-substituted protein has been prepared (Van Duyne *et al.* 1993) for use in multiwavelength anomalous diffraction (MAD) experiments (Hendrickson 1991) and has yielded data-quality crystals. A MAD intensity data set has been collected using beamline 1-5 at SSRL (Table 3.1).

Patterson map analysis of the MAD data using SOLVE (Terwilliger *et al.* 1999) has located four of the five possible selenium sites and shown the space group to be $I2_12_12_1$, rather than the far more commonly observed space group $I222$. Initial electron-density maps show clear protein and solvent regions. Solvent flattening in PHASES

(Furey *et al.* 1997) has significantly improved the quality of these maps and model building is under way.

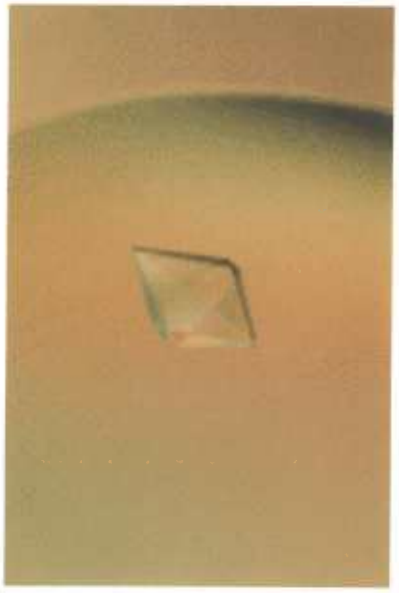

Figure 3.1 A crystal of MtaN. This crystal is approximately 0.8 mm long.

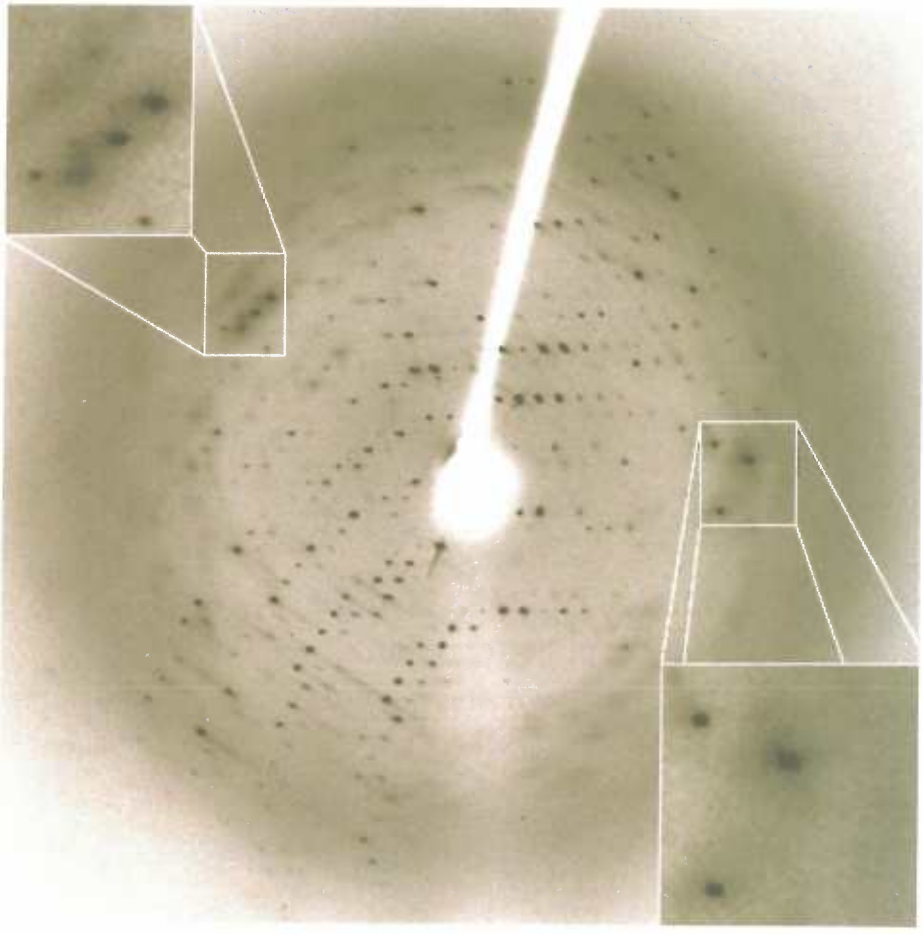

Figure 3.2 A diffraction image of MtaN. Insets show a 3-fold magnification of diffuse scattering in the 4.0-3.7 Å resolution shell.

Table 3.1
Data statistics.

Crystal	Native	Selenomethionine-substituted			
SSRL[a] Beamline	7-1	1-5			
Unit cell parameters (Å)	a = 49.4	a = 50.1			
	b = 67.8	b = 67.6			
	c = 115.0	c = 116.1			
Resolution limits (Å)	23.30 - 2.75	58.72 - 2.90			
Outer shell (Å)	2.82 - 2.75	2.98 - 2.90			
Wavelength (Å)	1.08	0.9226	0.97945	0.97988	1.06883
Observed reflections	42297	18,361	18,945	19,064	18,568
Unique reflections	5199	4502	4515	4523	4497
% Completeness (outer shell)	99.3 (99.0)	97.0 (100)	97.3 (100)	97.4 (100)	96.9 (93.5)
Mean $I/\sigma I$ (outer shell)	7.9 (2.7)	4.2 (1.4)	4.0 (1.5)	6.4 (1.6)	5.1 (2.0)
R_{sym} (%)[b] (outer shell)	7.2 (27.1)	6.7 (51.0)	6.8 (51.2)	5.3 (48.6)	5.1 (38.1)

[a] Stanford Synchrotron Radiation Laboratory

[b] $R_{sym} = \Sigma \left| I_o - I_{avg} \right| \square \Sigma I_o$

Chapter 4

CRYSTAL STRUCTURE OF MTAN, A GLOBAL MULTIDRUG-TRANSPORTER GENE ACTIVATOR

Michael H. Godsey[1], *Natalya N. Baranova*[2], *Alexander A. Neyfakh*[2] *and Richard G. Brennan*[1]

[1] Department of Biochemistry and Molecular Biology, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon, 97201-3098, U.S.A

[2] Center for Pharmaceutical Biotechnology, University of Illinois, 900 South Ashland Avenue, Chicago, Illinois, 60607, U.S.A.

Correspondence should be addressed to R. G. Brennan
E-mail: brennanr@ohsu.edu
Phone: (503) 494-4427
Fax: (503) 494-8393

Running Title: Crystal Structure of MtaN
JBC (2001). *Published as Papers in Press* October 1, 2001 as 10.1074/jbc.M105819200

Summary:

MtaN (Multidrug Transporter Activation, N-terminus) is a constitutive, transcriptionally active 109-residue truncation mutant, which contains only the N-terminal DNA-binding and dimerization domains of MerR family member Mta. The 2.75 Å resolution crystal structure of apo-MtaN reveals a winged-helix-turn-helix protein with a protruding 8-turn helix (α5) that is involved in dimerization by the formation of an antiparallel coiled-coil. The hydrophobic core and helices α1 through α4 are structurally homologous to MerR family member BmrR bound to DNA, whereas one wing (Wing 1) is shifted. Differences between the orientation of α5 with respect to the core, and the revolution of the antiparallel coiled-coil, lead to significantly altered conformations of MtaN and BmrR dimers. These shifts result in a conformation of MtaN that appears to be incompatible with the transcription activation mechanism of BmrR and suggest additional DNA-induced structural changes are necessary.

Introduction:

Bacterial multidrug resistance (MDR) is a growing threat to human health. One key component of MDR is the efflux of structurally and chemically diverse compounds, including antibiotics, antiseptics and disinfecants, by membrane-bound multidrug transporters (Paulsen *et al.* 1996; van Veen *et al.* 1999). Although often regulated by global regulators (Putman *et al.* 2000; Grkovic *et al.* 2001b) such as MarA (Rhee *et al.* 1998), which activates over a dozen genes (Alekshun *et al.* 1997; Barbosa *et al.* 2000), many MDR genes are regulated specifically, such as *qacA* by QacR (Grkovic *et al.* 1998) and *emrAB* by EmrR (Lomovskaya *et al.* 1995). In *Bacillus subtilis*, BmrR (Ahmed *et al.* 1994) and BltR (Ahmed *et al.* 1995), members of the MerR family (Summers 1992),

regulate transcription of the MDR-transporter genes *bmr* and *blt*, respectively. MtaN (multidrug transporter activation, N-terminus), another MerR family member, is a global activator of *B. subtilis* multidrug transporter genes, and constitutively activates transcription of *bmr* and *blt*, another putative membrane protein gene (*ydfK*) and its own gene (Baranova *et al.* 1999).

MerR proteins range from relatively small size, such as the *E. coli* MerR (144 residues per monomer) and *E. coli* ZntR (141 residues), to those over a hundred amino acid residues longer including *B. subtilis* BmrR (278 residues) or *Streptomyces lividans* TipAL (253 residues). These proteins form homodimers that regulate genes to combat a variety of cellular stresses. ZntR (Outten *et al.* 1999), CueR (Outten *et al.* 2000), PMTR (Noll *et al.* 1998) and MerR (Summers 1992) bind divalent metal ions to activate their respective metal resistance systems, while SoxR responds to oxidative stress through redox disassembly of its iron-sulfur centers (Hidalgo *et al.* 1997b). NolA is involved in the nodulation process in *Bradyrhizobium japonicum* by responding to nodulation factors from soybeans (Sadowsky *et al.* 1991). BmrR binds toxic lipophilic cations, although physiologically relevant ligand(s) of BmrR have yet to be identified (Ahmed *et al.* 1994). TipAL covalently binds the large antibiotic thiostrepton (Holmes *et al.* 1993). While MtaN is able to activate transcription of multidrug transporters and full-length Mta is closely related (40% sequence identity) to TipAL, Mta is not induced by thiostrepton (Baranova *et al.* 1999) and the ligand or ligands of Mta remain unknown.

The N-terminal "domain" of each MerR subunit, the most conserved segment, contains a winged helix-turn-helix motif (Shewchuk *et al.* 1989a) and the dimerization region, which comprises half of an antiparallel coiled-coil (Caguiat *et al.* 1999). This ~110 residue domain is the signature of the MerR family and it is likely to be structurally and functionally conserved. Beyond the winged helix-turn-helix motif, there appears to be no significant sequence or structural homology between MerR family members and other known gene regulators. The variable-length C-terminal domain of MerR proteins contains ligand or coactivator binding elements that have been tailored to recognize their widely divergent and non-overlapping signals. Not surprisingly, the larger proteins bind larger coactivators, while the smaller proteins appear to be the minimum size necessary to respond to a divalent cation.

The function of the C-terminus is to modulate the transcriptional activation of MerR family members by keeping the protein/DNA complex in a transcriptionally inactive form until a coactivator is bound, at which time repression is relieved and the protein is able to up-regulate transcription (Baranova *et al.* 1999; Godsey *et al.* 2000). MtaN is an unusual MerR family member because the protein lacks this modulation domain, which leads to its constitutive activation of cognate promoters (Baranova *et al.* 1999). Because MtaN constitutively activates its own transcription, cells containing *mtaN* produce high levels of this protein through positive feedback. Eventually, elevated levels of MtaN overcome its lower affinities for the *bmr* and *blt* promoters and those genes are activated (Baranova *et al.* 1999). MtaN appears to represent the smallest active form of the MerR family of transcriptional regulators.

An unusual feature of the genes that are regulated by MerR family members is the 19-base pair (bp) separation of the –10 and –35 promoter elements (Parkhill *et al.* 1990), which is 17-bp in most bacterial promoters (McClure 1985; Helmann 1995). The 19-bp spacer appears to prevent open complex formation by RNA polymerase (RNAP) in the absence of an activator (Parkhill *et al.* 1990). This unusual promoter structure has led to a model of transcription regulation by these proteins in which activation is achieved by DNA distortion and untwisting (Ansari *et al.* 1992; Summers 1992). The recent crystal structure of BmrR bound to DNA and coactivator has delineated a significant portion of the activation mechanism (Zheleznova-Heldwein *et al.* 2001). The ternary complex shows the center of the DNA-binding site is bent, untwisted and bunched-up, shortening the effective length of the DNA and reconfiguring the RNAP binding sites to resemble more closely a 17-bp spacer and allow open complex formation.

The BmrR-drug-DNA complex provides insight into one facet of transcription regulation by the MerR family. However, the extent of the conformational changes of these proteins needed to effect DNA-binding and transcription activation, if any, are unknown. To address this aspect of the mechanism of MerR family transcription activation, we solved the crystal structure of MtaN to 2.75 Å resolution. Comparison of the structures of MtaN and DNA/drug-bound BmrR reveals their overall structural similarity, as well as significant tertiary and quaternary differences.

Methods:

Data collection and phase determination

MtaN was expressed, purified, and crystallized as previously reported (Godsey *et al.* 2000). Both multiwavelength anomalous diffraction (MAD) and native X-ray intensity data were collected on cryocooled crystals at -170°C at the Stanford Synchrotron Radiation Laboratory (SSRL) on beamlines 1-5 and 7-1, respectively (Godsey *et al.* 2000). Native intensity data were collected at λ = 1.08 Å. MAD data were collected from a selenomethionine-containing protein crystal at four wavelengths (Table 4.1). Data were processed using MOSFLM (Leslie 1992). The structure of MtaN was determined by MAD phasing (Hendrickson 1991) as a special case of multiple isomorphous replacement (Ramakrishnan *et al.* 1997). Three of five possible selenomethionine sites were located using SOLVE (Terwilliger *et al.* 1999) and these revealed the space group to be $I2_12_12_1$. A fourth selenium site was located through difference Fourier analysis (Godsey *et al.* 2000). The fifth possible site, the N-terminal methionine, was not found and likely cleaved during normal protein processing by *E. coli*. The four sites were refined and electron density maps calculated after density modification, including solvent flipping (63% estimated solvent) and histogram matching as implemented in CNS (Brunger *et al.* 1998). The figure of merit for these phases increased from 0.69 in the initial MAD derived phases to 0.98 after density modification (resolution limit 58.0 Å – 2.9 Å). Initial electron density maps, calculated to 3.0 Å resolution, revealed a mostly α-helical structure, sections of well-connected backbone, and the locations of many side chains. The selenomethionine sites were later used to unambiguously determine the register of the protein chain.

Model building and refinement

An initial polyalanine trace was built manually to these maps with O (Jones *et al.* 1991). Iterative cycles of model building, including some side chains, with O and refinement with CNS continued until the usefulness of the MAD data was exhausted, at which point the MAD-derived model was used to "solve" the higher-resolution native intensity data set by molecular replacement using EPMR (Kissinger *et al.* 1999). Refinement and model rebuilding were done alternately to fit the complete model using sigma-A-weighted 2Fo-Fc and Fo-Fc maps (Figure 4.6). Prior to refinement, 5% of the native data was set aside for cross-validation and R_{free} was used as a measure of the validity and progress of the model. Residues 2-107 are included in the final model but the final 2 residues of the protein, a 2 residue linker, and the hexa-His tag are disordered and not included. The final R_{factor} and R_{free} are 0.227 and 0.287, respectively for data from 25.0 Å – 2.75 Å resolution. Four amino acids (residues Glu 67, His 71, Asn 73, and Lys 90) were refined as alanines because their side chain densities are missing or poor. The final model contains 16 solvent molecules. The stereochemistry of this structure was assessed with PROCHECK (Laskowski *et al.* 1993), which calculated 89% of all residues in the most favored region of the Ramachandran plot, and none in the generously allowed or disallowed regions.

Alignment

An alignment of the amino acid sequences of ten members of the MerR family was performed using CLUSTALW (Thompson *et al.* 1994) and aided by the structures of MtaN and BmrR. This alignment was examined using the computer programs Alscript (Barton 1993) and AMAS (Livingstone *et al.* 1993). Functional conservation

across the family was determined using a threshold value of 0.4 as described in the documentation of AMAS. Residues from the N-terminus through residue 69 of MtaN were aligned simply and consistently, however when different parameters were entered into the CLUSTALW program, alignments C-terminal to residue 69 varied slightly and those presented in Figure 4.3 represent its best fit. The alignment of BmrR to MtaN was confirmed by visual inspection of their structures.

The coordinates and structure factor amplitudes have been deposited in the RCSB Protein Data Bank (accession code 1JBG).

Results and Discussion:

Overall structure

The asymmetric unit contains a monomer of MtaN (Figure 4.1). The electron density is clear for 106 residues; the C-terminal 10 amino acyl residues (including the hexa-His-tag) are missing, as is the N-terminal methionine, which was not seen in either the selenomethionine-substituted or native structures. The side chains of residues Glu 67, His 71, Asn 73 and Lys 90 are disordered and have been modeled as alanines.

The topological arrangement of secondary structural elements of MtaN begins with β1 (residues 2-4), followed by α1 (residues 5-12) and α2 (residues 16-24), which are connected by a 3-residue turn and comprise the conserved and predicted helix-turn-helix motif (Shewchuk *et al.* 1989a). Residues 25-31 form a loop that connects α2 to strand β2 (residues 32-34), which is followed by a type II β-hairpin turn (residues 34-37) and strand β3 (residues 38-41). β3 is the center strand of a 3-stranded antiparallel β-sheet. A dipeptide connects β3 to helix α3 (residues 43-58), which leads into a tight

3-residue turn that connects to helix α4 (residues 62-70). A poorly structured loop (residues 71-75) connects the body (β1 through α4) to a protruding 8-turn helix α5 (residues 76-104). Thus, the topology of the MtaN monomer (β1-α1-α2-β2-β3-α3-α4-α5) is the same as that of the N-terminus of BmrR (Zheleznova-Heldwein *et al.* 2001).

The monomer contains two functional domains: the N-terminal DNA-contacting domain from β1 to α4 and the dimerization domain consisting of helix α5 (Figure 4.2). The DNA-binding domain is a member of the winged-helix-turn-helix family of proteins (Gajiwala *et al.* 2000a), consisting of a four-helix bundle and a three-stranded antiparallel β-sheet. The dimerization domain consists of the 8-turn α5 helix that forms a two-helix antiparallel coiled-coil with the other subunit. In BmrR α5 contains three additional turns of helix that extend into the C-terminal domain. The strong conservation of this fold and that described for BmrR (28% sequence identity) confirms the hypothesis that this structure would be general for the MerR family.

DNA-binding domain – the body

The structure of MtaN is stabilized by a hydrophobic core, which consists of side chains from α1 (Val 5, Val 8, Ala 9), α2 (Leu 19, Tyr 22, Asp 23), β3 (Arg 39, Tyr 41), α3 (Leu 46, Leu 49, Ile 52, Phe 55, Ile 58), and α4 (Leu 62, Ile 65, Met 68, Leu 69) and turns between α1 and α2 (Val 14), α2 and β2 (Ile 25, Leu 27, Leu 28, Pro 30), and α3 and α4 (Phe 60). All but two of these 23 core residues (Ile 25 and Pro30) are well conserved across the MerR family (Figure 4.3).

In addition to the hydrophobic component of the core, Asp 23 and Arg 39 form a buried salt bridge. This salt bridge buttresses the position of β3. In BmrR, this arginine (BmrR:Arg 43) is turned away from the carboxylate group of the aspartate (BmrR:Asp 26) to contact the DNA phosphate backbone. An Arg is absolutely conserved at this position across the MerR family and the Asp is either an Asp or Glu in all members but MerR, where it is a Gln. While a formal possibility is that the Asp 23-Arg 39 salt bridge is in equilibrium with an unbridged conformer, such as that seen in BmrR, the high ionic strength of the MtaN crystallization conditions (up to 5.0 M LiCl) would be expected to disfavor the formation of this observed salt bridge strongly. Given that, Asp 23 and Arg 39 are found to interact in this environment indicates that this is a stable and physiologically relevant interaction. Thus this salt bridge interaction, and its subsequent DNA-induced breaking, is likely to be conserved in all MerR family members.

The reason for sequence conservation of MtaN and BmrR is clear. Of the body, defined as β1 through α4, 26 of 69 MtaN and BmrR residues are identical, and of those 26, 15 are found in the core, and an additional 5 in turns. When conservative substitutions are included, the number of homologous residues rises to 36 and of those, 21 are found in the core. Thus, the observed sequence conservation between the two proteins ensures the structural conservation of this hydrophobic core. The same reasoning appears to apply across the whole family as these residues are among the most conserved in the N-terminal domain. Our analysis leads to the conclusion that DNA binding by MerR proteins does not significantly alter the structure of the hydrophobic core of the DNA-binding domains.

An overlay of α–carbons of helices α1 through α4 of MtaN onto the corresponding BmrR atoms results in a root mean square (r.m.s.) deviation of 0.75 Å. That these four helices overlay so well suggests that either this domain of MtaN has taken the DNA-bound formation even in the absence of DNA, or more likely there is no difference in relative positions of these helices between the DNA-bound and free forms. However, residues Asp 47 and Ser 48 of BmrR α3 are displaced from their MtaN counterparts (Asp 43 and Ala 44) by their connection to the β-sheet, which takes a different conformation, and by their direct interaction with the drug-binding domain of BmrR, which is not present in MtaN.

While the cores of MtaN and BmrR and the positions of the body helices (α1-α4) are the same, a structural difference is evident in the position of Wing 1 (β2-turn-β3). Specifically, MtaN displays a type II β-turn (Thr 34-Gly 37), while BmrR does not contain this classic hairpin because of a single residue insertion in this area, and is thus better described as a small loop. In addition, MtaN Wing 1 makes crystal lattice contacts, whereby Tyr 38 stacks against a symmetry-related Tyr 38. In BmrR this residue (BmrR:Tyr 42) interacts with a base in the minor groove. As a result, the Cα of Asp 35 (BmrR:Asp 39), which is located at the crux of the β-turn, moves 8.3 Å. Either interaction (protein-protein or protein-DNA) might be enough to displace the end of this β-sheet and therefore, Wing 1. Such wing flexibility is well documented in other winged-helix proteins (Jin *et al.* 1999a; Jin *et al.* 1999b). Regardless of its position in the current structure, Tyr 38 is likely to be involved in DNA binding by MtaN as well.

The more global shift in the rest of the wing is more likely due to the absence of DNA because of the different interactions of conserved residue Arg 39, which forms a salt bridge to Asp 23 in MtaN and a DNA backbone contact in BmrR (BmrR:Arg 43).

Dimerization domain

The MtaN dimer is stabilized primarily by the formation of an antiparallel coiled-coil between the amphipathic α5 helices. Coiled-coils are characterized by a heptad repeat $(\boldsymbol{abcdefg})_n$, in which the ***a*** and ***d*** positions are typically occupied by hydrophobic residues and form the interface between the interacting helices (Nishikawa *et al.* 1976; Fujinaga *et al.* 1993). In MtaN, the hydrophobic core of the interface consists of the side chains of residues Leu 80 (***d***1), Leu 87 (***d***2), Met 94 (***d***3), Ile 98 (***a***3), Ile 101 (***d***4) and Leu 105 (***a***4), and the methylene carbons of Lys 84 (***a***1) and Lys 91 (***a***2). In the antiparallel conformation found in MtaN, van der Waals contacts are from ***d***1 to ***a***4', ***a***1 to ***d***4', ***d***2 to ***a***3', ***a***2 to ***d***3', (where ' indicates the dimer partner) as well as the symmetry imposed interactions (Figure 4.4). In addition to forming the antiparallel coiled-coil, the C-terminal end of the α5 helix also interacts with the C-terminus of α3'. Contacts are found between the side chain of Phe 54 and Cγ of Thr 104', and the alkyl side chains of Ile 58 and Ile 101'. van der Waals contacts between Glu 57 and Met 97' complete the dimerization interface. Dimerization buries 738 $Å^2$ of accessible surface area per monomer, which is average for many oligomeric proteins (Conte *et al.* 1999).

Beyond the hydrophobic interactions, two ***a*** residues of MtaN, Lys 84 and Lys 91 and their dyadic mates, form salt bridges to Asp 102' and Asp 95' respectively,

while the corresponding BmrR residues do not. Inter-helix ionic interactions are common among both parallel and antiparallel coiled-coils and serve to stabilize the dimer and prevent unwanted heterodimerization (Monera *et al.* 1994; Zhou *et al.* 1994).

An antiparallel coiled-coil was first indicated in MerR (Lupas 1997; Wolf *et al.* 1997; Zeng *et al.* 1998) and predicted to occur in all MerR family members (Caguiat *et al.* 1999), which was confirmed by the structure determination of BmrR. In both MtaN and BmrR, all ***d*** positions are occupied by hydrophobic residues, while their ***a*** positions vary significantly (Figure 4.4). Specifically, at MtaN ***a*** positions Lys 91, Ile 98 and Leu 105, the respective BmrR residues are Leu, Glu and Lys. Where both proteins have hydrophilic ***a*** residues they are oppositely charged (MtaN:Lys 84, BmrR:Glu 88). Overall, the buried residues of the coiled-coil (80 through 105) are only partially conserved across the family, even between MtaN and TipAL, the most closely related MerR protein (Baranova *et al.* 1999) (49% sequence identity). Only residues corresponding to Ile 101, which is always hydrophobic, and Leu 80, Leu 87, Met 94 and Ile 98, which are usually hydrophobic, are reasonably conserved (Figure 4.3). Thus, the variation of buried residues serves to stabilize the dimer and likely contributes to the prevention of heterodimerization.

Conformational differences between two MerR family members

An overlay of the conserved four-helix core of one monomer of MtaN onto the corresponding core of DNA-bound BmrR revealed a significant shift in the relative positions of the recognition helix (α2') of the other subunit (Figure 4.5). In MtaN the center-to-center distance of these helices is 33.3 Å, close to the 34 Å repeat distance of

canonical B-form DNA and consistent with their binding to consecutive major grooves. In the transcription-activated conformation of BmrR these helices are only 30.6 Å apart, which corresponds to their major-groove binding to a shortened and undertwisted DNA double helix (Zheleznova-Heldwein *et al.* 2001). In addition to the distances, the relative positions of these helices have changed with respect to each other. The resulting position of the MtaN α2' is offset from the BmrR α2' by 7.5 Å, largely due to the lateral twist of 15° of the dimer partner, rather than a simple direct lengthening between the major groove binding helices (Figure 4.5 *a*, *b*).

The rotation between subunits is the result of two conformational changes that occur in the antiparallel coiled-coil. When MtaN and BmrR helices α1 through α4 are overlaid, a shift in the relative positions of their α5 helices is evident (Figure 4.5*a*, *b*). In comparison to MtaN α5, the BmrR α5 has rotated approximately 6.5° up and away. The body of the dimer partner moves to match this relocation to maintain the contacts between helices α5 and α3'. This relatively small rotation is doubled by the same rotation of the dimer partner, and further amplified by the length of the coiled-coil. In addition to the rotation of the α5 helix, the relative conformations of the antiparallel coiled-coils of MtaN and BmrR are different. When α5 of MtaN is overlaid onto the α5 of BmrR (r.m.s. deviation = 0.63 Å for the α-carbons), the α5' helices do not overlay (Figure 4.5*c*, *d*). Rather, the C-terminal end of the MtaN α5' helix has revolved ~15° in a counter-clockwise direction around the overlaid α5 helices. The movement of the N-terminus of α5' helix is smaller, but changes the direction of the helical axis to match the revolution that occurs at the C-terminal end. This revolution rotates the body, i.e.,

the DNA-binding domain, of the dimer partner around the axis of the coiled-coil and swings it towards the other body domain, thereby accounting for the observed expansion of the recognition helices of MtaN.

DNA-induced conformational changes

MtaN is a constitutive activator, yet the dimer structures of MtaN and BmrR, the latter of which is in its transcription-activated conformation, are different. Perhaps the differences reflect dissimilar DNA-binding modes in which MtaN twists its promoter DNA to a lesser degree. Alternatively, the DNA binding site might play a role in the induction of additional conformational changes in MtaN so that it more closely resembles BmrR.

DNA-docking experiments reveal that the MtaN dimer is unable to bind the BmrR-activated DNA (Zheleznova-Heldwein *et al.* 2001) because its α2 major groove recognition helices are too far apart and in the wrong orientation to fit into the major grooves. The solved crystal form of MtaN is also unable to bind canonical B-form DNA because the α2 helices are tilted incorrectly to fit directly into adjacent major grooves and Wing 1 clashes with the DNA backbone. Thus at the least, MtaN requires minor structural adjustment in the twist of its α2 helices and more significant changes in the position of its β-sheet (Wing 1) in order to bind either DNA conformation

Given the results of our docking experiments a binding and activation mechanism can be envisioned. In this proposal the first step is MtaN binding to a B-like DNA conformation. This would likely be concomitant with or followed by the breaking of the Asp 23-Arg 39 salt bridge. The disruption would allow Arg 39 to

contact the DNA phosphate backbone, perhaps as observed in the BmrR-*bmr* promoter complex (Zheleznova-Heldwein *et al.* 2001) and remove a key constraint that holds MtaN in a non-activating conformation. Additional structural changes would be transmitted through the coiled-coil and allow the MtaN conformation to maximize its DNA contacts. This in turn could elicit DNA conformational changes, which would result in an activated conformation of the MtaN-*mta* promoter complex that may more closely resemble that of the BmrR-*bmr* complex. The structure of an MtaN–*mta* promoter complex should provide more understanding of the DNA-binding and transcription activation mechanisms of this MerR family member.

Acknowledgements:

The authors wish to thank E. E. Zheleznova-Heldwein for sharing the BmrR coordinates with us prior to publication. This work was supported by grants from the NIH (AI 48593 to R.G.B. and training grant GM08617-05 to M.H.G.), NSF (MCB-9816983 to A.A.N.) and the N. L. Tartar trust. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

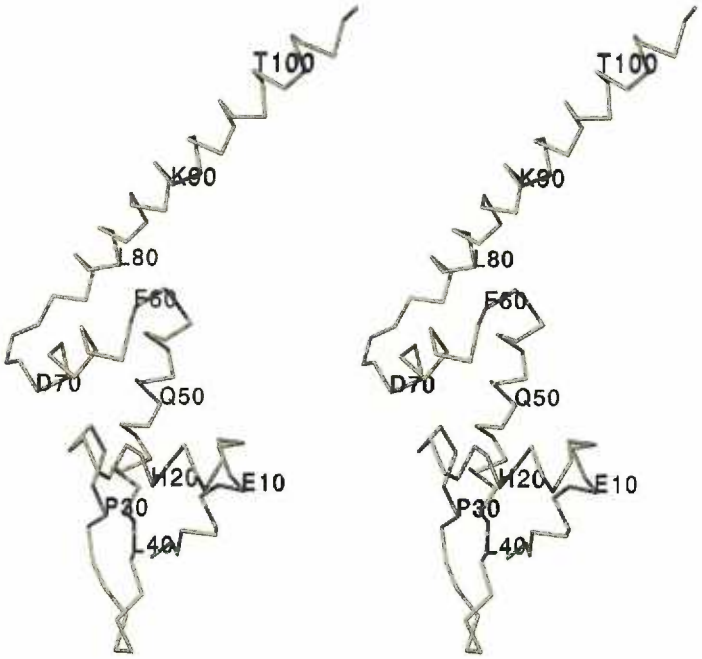

Fig 4.1. A stereo view of the overall architecture of MtaN. Every tenth residue is labeled. Figures 1, 2, 5 and 6 produced with Swiss PDB viewer (Guex *et al.* 1997) and POV-ray (www.povray.org).

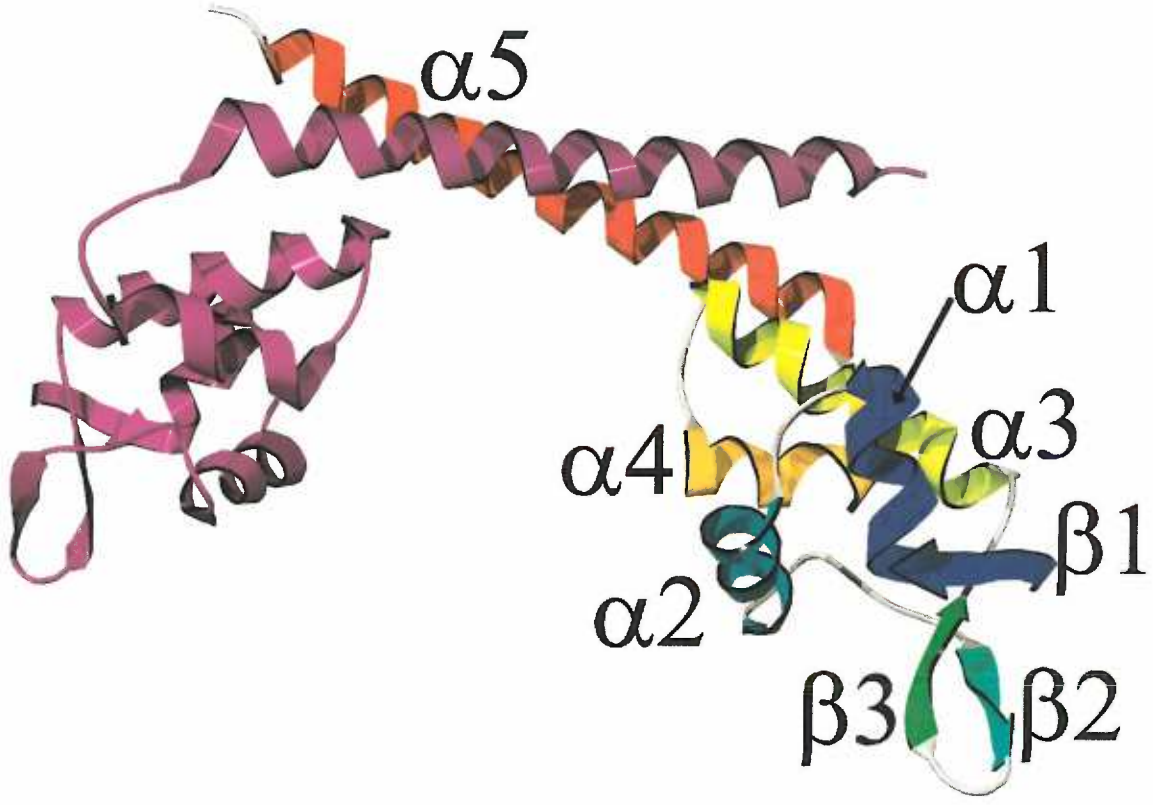

Fig 4.2. A ribbon diagram of the MtaN dimer. One subunit is colored from blue at the N-terminus to red at the C-terminus. The other is colored purple. Secondary structural elements are labeled. The "body" domain includes β1, α1, α2, β2, β3, α3 and α4. The α5 helices form the antiparallel coiled-coil.

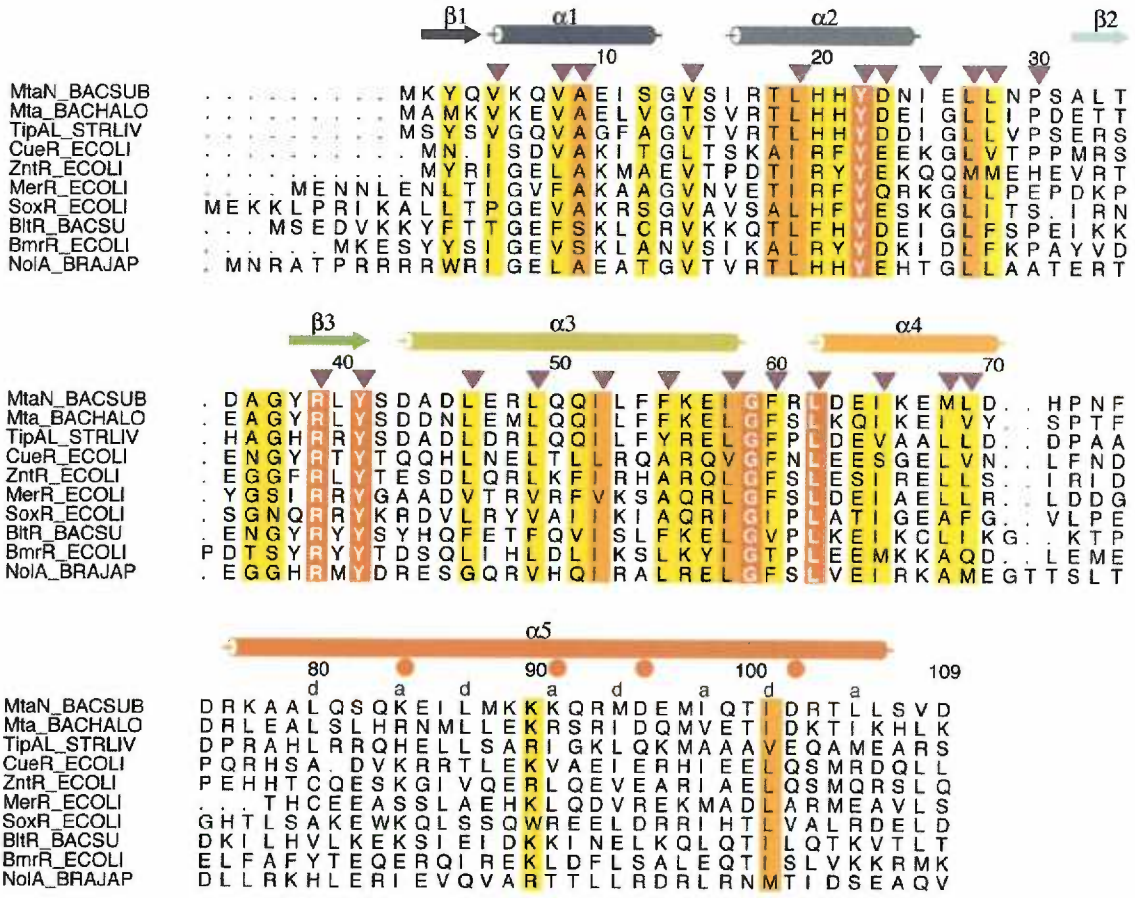

Fig 4.3. A sequence alignment of MtaN and other MerR family members. Conserved residues are color-coded, with decreasing conservation from red shading (absolute) to white (minimal). The secondary structure elements of MtaN are shown above the alignment and colored as in Figure 4.2. Purple arrowheads denote residues with side chains contributing to the core. Filled red circles are MtaN residues involved in intermolecular salt-bridges. The ***a*** and ***d*** positions of the antiparallel coiled-coil are indicated.

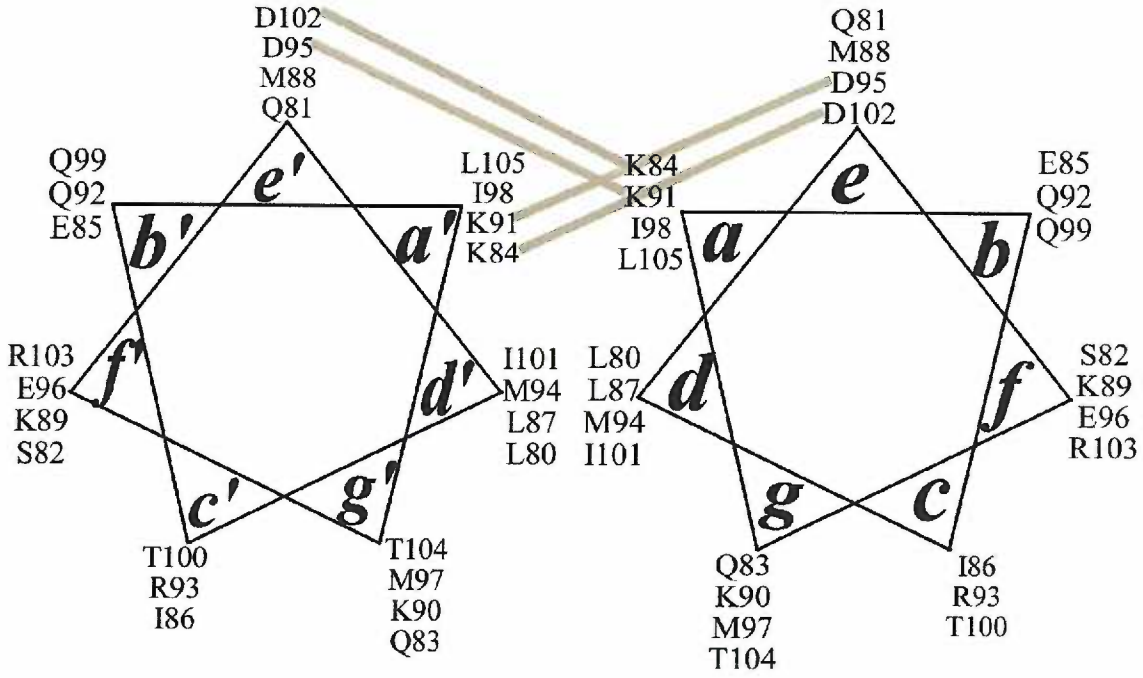

Fig 4.4. A schematic diagram of the antiparallel coiled-coil of MtaN. Inter-subunit salt bridges between the ***a*** and ***e*** positions are depicted by gray lines. Residues at the ***a*** and ***d*** positions form the hydrophobic core of the interacting helices.

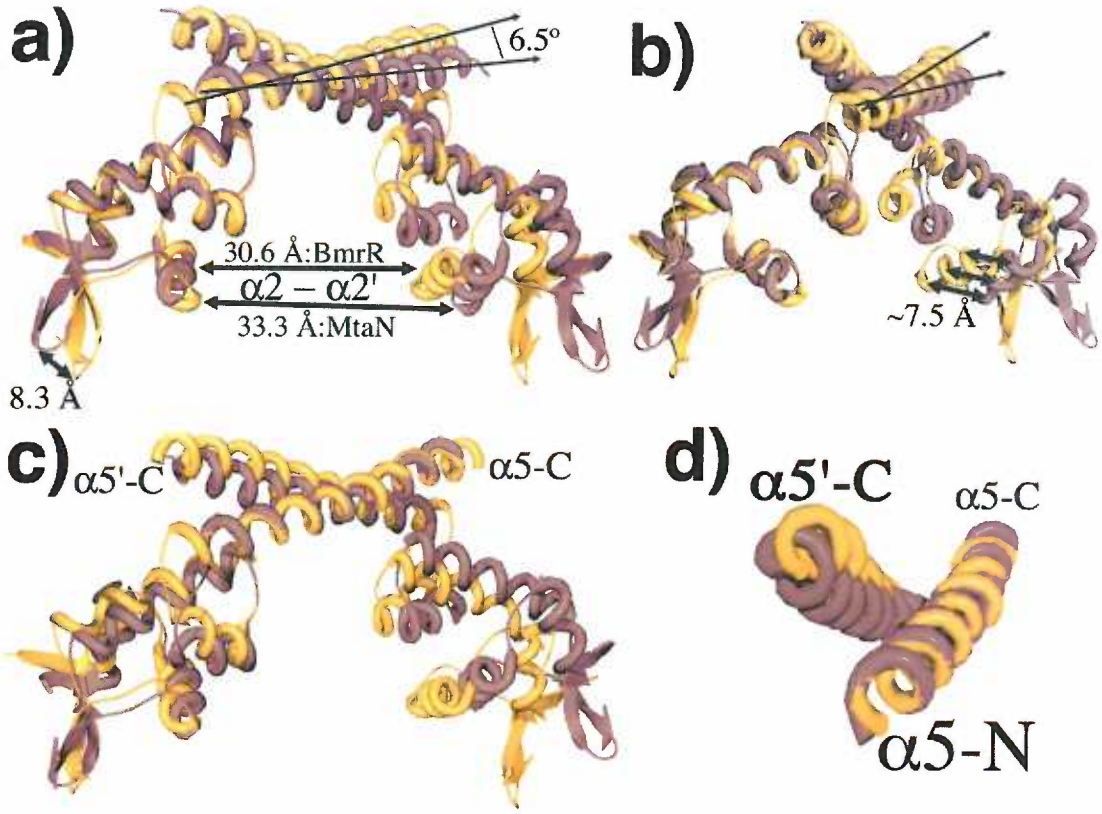

Fig 4.5. Conformational differences between MtaN and BmrR. MtaN is in purple and the first 110 residues of BmrR are in gold. *a*, Superimposition of helices α1 through α4 (left monomers) reveals the rotation of the α5 helix, altered positions of the β-sheets and different orientations of the α2 and α2' helices, which in MtaN are farther apart and rotated counterclockwise. *b*, A view of the same overlay in *a* rotated by ~45°. This view highlights the different locations of the MtaN and BmrR α2' helices, which are shifted ~7.5 Å. *c*, Superimposition of the α5 helices. The α5' C-terminal end of BmrR is revolved around the overlaid α5 helices as compared to MtaN, resulting in a large shift in the body of the dimer partner and a smaller effect on the position of the monomer body. *d*, the same overlay as *c*, but looking at the N-terminus of α5 and C-terminus of α5'. The remainder of the protein has been deleted for clarity.

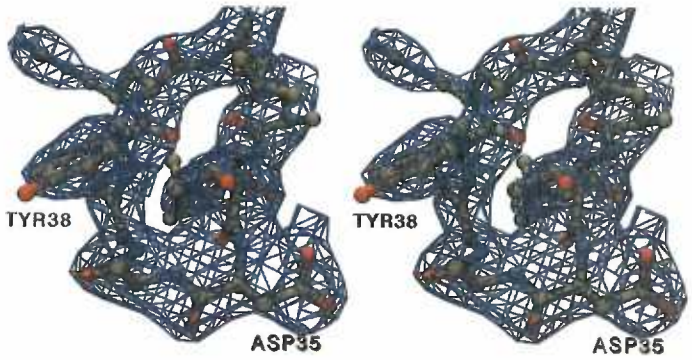

Fig 4.6. A stereo view of sigma-A-weighted simulated annealing (2Fo-Fc) omit map electron density in the area of the β-turn between β2 and β3, contoured at 1σ.

Table 4.1

Data and refinement statistics.

Data Collection and MAD phasing statistics					
Crystal	Native	Selenomethionine-substituted			
SSRL[a] Beamline	7-1	1-5			
Unit cell parameters (Å)	a = 49.4	a = 50.1			
	b = 67.8	b = 67.6			
	c = 115.0	c = 116.1			
Resolution limits (Å)	23.30 - 2.75	58.72 - 2.90			
Outer shell (Å)	2.82 - 2.75	2.98 - 2.90			
Wavelength (Å)	1.08	0.9226	0.97945	0.97988	1.06883
Observed reflections	42297	18,361	18,945	19,064	18,568
Unique reflections	5199	4502	4515	4523	4497
% Completeness (outer shell)	99.3 (99.0)	97.0 (100)	97.3 (100)	97.4 (100)	96.9(93.5)
Mean $I/\sigma I$ (outer shell)	7.9 (2.7)	4.2 (1.4)	4.0 (1.5)	6.4 (1.6)	5.1 (2.0)
R_{sym} (%)[b] (outer shell)	7.2 (27.1)	6.7 (51.0)	6.8 (51.2)	5.3 (48.6)	5.1(38.1)
MAD Figure of merit			0.688		

Refinement statistics (native data set)	
Number of reflections (working/test)	4551/261
Number of nonhydrogen protein atoms	861
Solvent atoms	16
Resolution (Å)	25.0 - 2.75
R_{cryst} / R_{free} (%)	22.8 / 28.7
Bond length deviation (Å)	0.013
Bond angle deviation (°)	1.5
Average B-factor (Å^2)	81.61

[a] Stanford Synchrotron Radiation Laboratory

[b] $R_{sym} = \Sigma \left| I_o - I_{avg} \right| \square \Sigma I_o$

Chapter 5

BINDING OF MTAN TO THE MERR FAMILY PROMOTERS

Introduction

MerR proteins bind to an unusual promoter structure with a 19 base-pair (bp) spacer between the –10 and –35 elements, in contrast to the more commonly found 17 bp (McClure 1985; Helmann 1995). This was shown to be vital to the function of MerR in a series of experiments where this spacer was altered by site-directed mutagenesis and the effects on binding and activation were determined (Parkhill *et al.* 1990). Effects were measured *in vitro* by gel-shift binding assay and *in vivo* transcriptional activation. Shortening of the spacer to 18 or 17 bp led to a highly active promoter, even in the absence of MerR. Lengthening of the spacer to 20 or 21 bp prevented transcriptional activation by MerR (Parkhill *et al.* 1990). The gel-shift binding assay revealed that MerR bound the 18 or 20 bp spacers with slightly increased association constants, but that binding to the 17 or 21 bp spacers was significantly weaker than wild-type. These results suggest that the 19 bp spacer is required for correct induction by MerR, but that binding by MerR is more tolerant of alterations in this spacing. A similar study was done with SoxR, which also showed that correct functioning of the protein depended upon the 19 bp spacer (Hidalgo *et al.* 1997a). MtaN also binds to promoters with the MerR-type 19 bp spacer.

MtaN was discovered, as described in chapter 1, as a mutant that increased resistance to ethidium bromide (Baranova *et al.* 1999). This resistance is mediated

through the action of the Bmr and/or Blt multidrug pumps, and MtaN was found to activate transcription of the genes for those pumps. MtaN was subsequently shown by DNase I footprinting and *in vitro* transcription assays to bind the MerR-type promoters for those genes with lower affinity than the promoter of its own gene. The estimated relative concentrations of MtaN that were required to generate a footprint were 1:3:9 for *mta*:*bmr*:*blt,* although values for the binding affinity were not determined (Baranova *et al.* 1999).

Based upon the observed ability of MtaN to activate *bmr* and *blt* in addition to its own gene, and the disparity of the sequences of these promoters, Neyfakh and colleagues hypothesized that MtaN could activate other promoters of the MerR family (Baranova *et al.* 1999). To test this, a computerized search for other possible MtaN binding sites in the *B. subtilis* genome was undertaken using the similarities of the *bmr*, *blt* and *mta* promoters as the search criteria. A possible binding site was found upstream of a hypothetical membrane protein gene, *ydfK*. Northern hybridization found increased levels of *ydfK* mRNA in cells with MtaN, supporting the assertion that MtaN was also activating this gene. However, *in vitro* binding experiments with MtaN and the *ydfK* promoter were not conducted. YdfK is predicted be a membrane protein with seven TM domains, in contrast to the 12 TM domains of Bmr and Blt. No studies of YdfK have been done to suggest that it is a multidrug transporter. BLAST searches (Altschul *et al.* 1990; Altschul *et al.* 1997) of the databases with YdfK reveal significant homology only to other hypothetical proteins. Nothing else has been published about the possible role of YdfK to date.

The ability of MtaN to activate the transcription of several genes, with apparently similar binding affinities for the promoters, leads to the question of discrimination of binding by MtaN. The possibility that MtaN acts as a non-specific activator of all MerR promoters was raised because it was able to activate transcription at all of the MerR promoters examined (Baranova *et al.* 1999). If this is the case, and MtaN is a global activator, it would imply that Mta is a global regulator.

In studying MtaN, Baranova *et al.* also found that the full-length Mta was able to produce a footprint on its own promoter, but only at concentrations >27 times higher than those required for MtaN footprinting. Full-length Mta did not produce a footprint on the *bmr* or *blt* promoters, suggesting that the truncation of the protein not only increased its affinity for cognate DNA, but also reduced its discrimination for the *mta* promoter. While some MerR family proteins have been found to have higher affinity for specific DNA in the presence of their co-activators (TipAL (Holmes *et al.* 1993), BmrR (Ahmed *et al.* 1994)), others have shown similar affinities with and without co-activators (ZntR (Brocklehurst *et al.* 1999; Outten *et al.* 1999), MerR (Frantz *et al.* 1990; Heltzel *et al.* 1990)). This effect cannot be determined for Mta because the ligand of Mta remains unknown, although thiostrepton and known ligands of BmrR do not induce Mta (Baranova *et al.* 1999). However, removal of the C-terminal domain of Mta has the clear effect of increasing affinity for DNA and possibly reducing discrimination. This finding supports speculation that Mta might not be a global regulator, but that it normally has a specific role and is merely converted to a global activator in the absence of the C-terminal domain.

In the *B. subtilis* genome (Kunst *et al.* 1997), one gene adjacent to *mta* is a *ywnE*, a gene for a hypothetical protein with significant homology to cardiolipin synthases from several species. Cardiolipin, also known as phospholipid diphosphatidylglycerol, is a membrane phospholipid that is best known as being enriched in the inner mitochondrial membrane. Cardiolipin is found there probably because it is able to reduce the permeability of the inner membrane to protons, and therefore increase the proton motive force (PMF) across the membrane. Cardiolipin is also found in the membrane of *B. subtilis* (Filgueiras *et al.* 1980). The PMF is the energy source for the Bmr and Blt transporters. The effects of the *mta1* (MtaN) mutation on the PMF and the expression of *ywnE* have not been determined.

To characterize the DNA-binding properties of MtaN and the potential function as a global MerR family activator, fluorescence polarization (or anisotropy) (Lundblad *et al.* 1996) experiments were undertaken with purified MtaN and fluoresceinated oligonucleotides representing the DNA-binding sites of various MerR family member proteins (Table 5.1).

Materials and methods

Fluorescence polarization is based upon equilibrium binding of molecules in solution. Fluorescent molecules absorb and emit light at characteristic wavelengths. When the incident (absorbed) light is polarized, the emitted light will be preferentially polarized as well. The degree of polarization will depend upon the fluorescent lifetime and the motion of the molecule. Molecules with increased motion emit light that is less polarized due to tumbling and associated randomization of emission directions. It is this

fact that is used in fluorescence polarization studies. The rate of tumbling of a molecule depends upon, among other factors, the size of the molecule and the temperature. Upon binding of a fluorescently-labelled macromolecule to another macromolecule, the size of the complex will increase and the rate of tumbling will decrease. By slowing the motion of the macromolecule and associated fluorescent group, the polarization of the emitted light will increase. By measuring the polarization (or anisotropy) of the emitted light as a function of the concentration of macromolecules, the fraction of bound fluorescently-labelled molecule can be determined. Polarization and anisotropy are defined and related:

$$P = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + I_{\perp}), \quad A = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + 2I_{\perp}), \quad A = 2P/(3-P), \quad P = 3A/(2+A)$$

in which intensities of emitted light parallel ($I_{\parallel}$) and perpendicular ($I_{\perp}$) to incident light are measured. The polarization of fluoresceinated DNA was measured with increasing concentrations of purified protein, and a binding curve was fit to the data as described in the next paragraph.

MtaN was purified as previously described in chapters 3 and 4. Protein was taken directly from the dialysis tubing, and after the concentration of the "stock" was determined by A_{280}, it was diluted to appropriate concentrations (typically 2 μM, 10 μM, 50 μM, or 100 μM) in FA binding buffer (10 mM Tris, pH 7.6, 50 mM KCl, 5% glycerol). Protein was kept at room temperature (~24 C) throughout purification and binding experiments. Fluorescinated oligonucleotides were purchased from Oligos Etc. (Wilsonville, OR) or Sigma-Genosys (The Woodlands, TX) and stored dry at -20° C until annealing. One 5' fluorescein-labeled strand and the complementary unlabelled strand

were annealed by dissolving the oligonucleotides to a duplex concentration of 100 μM in FA binding buffer, followed by a 1:10 dilution into FA buffer to a 50 μl total volume. This solution was then heated to 95° C for ten minutes and allowed to cool slowly to room temperature. Concentration of the duplex was confirmed by reading the absorbance at 260 nm. This 10 μM duplex stock can be stored at -20° C, or diluted to 2 μM (working concentration) in FA binding buffer for immediate use. The oligonucleotides used were 33-mers unless otherwise noted, as this length extends beyond the footprint of MtaN on *mta* (Baranova *et al.* 1999) and had been found to bind well in preliminary studies (data not shown). The binding sites from the *mta*, *ydf*, *bmr* and *blt* promoters were chosen because they had been previously shown to be bound by MtaN. The *tip* promoter site was chosen because of the sequence homology between Mta and TipAL. The sites from the *E. coli znt* and *cue* promoters were tested, because they are bound by metal-binding MerR proteins. Their assay would allow a direct comparison of the affinity of MtaN for the drug-binding and non-drug binding MerR promoters. The *znt* site also has 20 bp spacer between the –10 and –35 promoter elements, which is unusually long even for MerR family binding sites. All of the above sites are known to be bound by MerR family member proteins. A modified PurR binding site, the *purF* operator, was used as a negative control.

Fluorescence polarization experiments were conducted using a PanVera Beacon 2000 instrument (Madison, WI). To a borosilicate tube, 1 mL of FA buffer was added and measured as the blank. Next, 1 μL of a 2 μM solution of the oligonucleotide to be tested was added to the tube, vortexed briefly and measured. Then 1 μL of a poly d(I-C)

stock (1 mg/ml) was added to the tube as a control for non-specific DNA binding, vortexed briefly and measured as the 0 nM protein data point. Finally, the binding curve is obtained by 1 μL additions of protein solution to the tube. Upon each addition the sample is vortexed lightly and polarization is measured after a 30 second incubation in the instrument. The 30 second incubation allows the binding to come to equilibrium. The concentration of the protein solution was varied to obtain meaningful results because of significant differences in the K_d values. To confirm thermal stability, temperature readings were taken periodically by inserting a thermometer immersed in 1mL H_2O into the measurement opening of the machine. These measurements never varied more than 1° C from 24° C during the course of these experiments.

Data were analyzed using the commercially available software packages KaleidaGraph and SigmaPlot. Data points were entered as measured polarization vs. MtaN protein concentration. Binding curves were fit to the following equation:

$$P_{measured} = \{(P_{max}\,[protein]) / (K_d + [protein])\} + P_0$$

by the curve-fitting algorithms in the software. $P_{measured}$ is the reading from the instrument. P_{max} is the maximum increase in polarization and P_0 is the polarization at zero protein concentration. P_{max}, K_d and P_0 are fit using least-squares regression analysis.

Results

MtaN was found to bind its own promoter with greater discrimination than was expected from the earlier reported results (Baranova *et al.* 1999). MtaN binds its (33-mer) cognate promoter with a K_d of 4.6 ± 0.8 nM (Table 5.2, Figure 5.1). MtaN

displayed a K_d of 6.1 ± 1.7 nM for a 30-mer oligonucleotide (shorter by one nucleotide on the 5 end and two nucleotides on the 3 end than the 33-mer) also containing the *mta* promoter site, showing that the length of the 33-mer duplex *mta* promoter was sufficient bind maximally. The K_d value obtained for the binding of MtaN to the *mta* promoter varied slightly from one protein purification to the next, presumably due to small differences in the amount of active protein in the preparation, and was found to range up to 11 nM. However, the error in these measurements generally overlapped indicating statistically insignificant differences between experiments. For comparison, Dr. Joy Huffman has been studying MerR family member PMTR, a metal-binding protein from *Proteus mirabilis* (Noll *et al.* 1998) and its interaction with its cognate promoter. She has found that PMTR has a K_d for its promoter of 20 ± 5 nM in the absence of Zn^{2+}, and 27 ± 4 nM in the presence of Zn^{2+} (Dr. Joy Huffman, personal communication). *E. coli* MerR has been reported to have a K_d for its promoter of 8 ± 5 nM (Parkhill *et al.* 1990). As an example of a high-affinity protein/DNA interaction not of the MerR family, the K_d of PurR for *purF* operator was reported as 2.4 ± 0.4 nM (Glasfeld *et al.* 1999). The affinity of MtaN for the *mta* promoter site is in this order of magnitude. Thus this high-affinity interaction is comparable to reported and unreported interactions of MerR family proteins with DNA, and PurR with the *purF* operator.

Other than the *mta* promoter, MtaN displayed the highest affinities for the *ydf* and *bmr* promoters with K_d values of 240 ± 40 nM and ~260 nM, respectively, approximately 60-fold higher than the value for the *mta* promoter (Figure 5.2). The values for the *bmr* and *blt* (below) promoters are averages obtained using oligonucleotides from two

different suppliers. The difference in K_d values of MtaN for the *mta* and *bmr* promoters from the different companies is statistically insignificant. While no attempt was made previously to quantify the binding affinity of MtaN for the *ydf* promoter, the *bmr* promoter was studied by DNase I footprinting (Baranova *et al.* 1999). This footprinting study found that approximately three times the concentration of MtaN was required to footprint the *bmr* promoter as was needed to footprint the *mta* promoter.

The *ydf* result is not surprising because there had been no previous evidence to compare the affinity of MtaN for this site to any other. That MtaN binds the *ydf* promoter and the *bmr* promoter with similar affinities suggests that if Mta is a global regulator then the *ydfK* gene is also part of its regulon. However, the *ydfK* gene is also located immediately downstream of *ydfL*, a gene predicted to encode a MerR family member protein. Based upon the structure of the other MerR family member operons in *B. subtilis*, this suggests that YdfL could specifically regulate *ydfK*. No work has been published on either *ydfL* or *ydfK* other than the reference to them in the Mta discovery paper (Baranova *et al.* 1999).

In our experiments where the dissociation constant approaches or exceeds 1 μM, the plot of polarization *versus* protein concentration becomes more linear, making precise K_d determination more difficult. Given this limitation, the observed K_d of MtaN for the *blt* promoter is ~2,600 μM, approximately 500-fold higher than for the *mta* promoter (Table 5.2). This contrasts with the footprinting studies, which had found that only nine times as much MtaN was required to footprint the *blt* promoter as was required to

footprint the *mta* promoter. The different values of K_d obtained for the *blt* oligonucleotides from the two suppliers may be, in part, due to differences in the preparations of oligonucleotides, although because the K_d is high, the errors are large and nearly overlap. It is now clear that MtaN displays greater discrimination for its own promoter than was expected over those of *bmr* and *blt*.

The four promoter sites described above were taken from *B. subtilis*. Because MtaN had been shown to be able to bind or activate the four promoters examined from *B. subtilis*, we chose to examine other bacterial MerR family member binding sites. The *tip* promoter from *Streptomyces lividans* was chosen because of the high sequence homology of Mta and TipAL. It should be noted again that thiostrepton, the ligand of TipAL, was not found to bind Mta (Baranova *et al.* 1999). MtaN bound to the *tip* promoter with a K_d of 1 μM, almost 170-fold higher than for its own promoter, and ~3-4 fold higher than for the *bmr* or *ydf* promoters (Figure 5.3). Thus, while the overall sequences of Mta and TipAL may be closely related, it appears that their cognate promoters are not (Table 5.1).

Binding sites for two metal-binding MerR family proteins, ZntR and CueR, were also examined. ZntR is activated by zinc (Brocklehurst *et al.* 1999; Outten *et al.* 1999) and CueR by copper (Outten *et al.* 2000). MtaN bound the *cue* and *znt* promoters with K_d values of 1 μM and 3.5 μM, respectively (Figure 5.3). The *E. coli znt* promoter contains a 20 bp spacer between the –10 and –35 promoter elements, unusually long even for MerR-type promoters. Again, K_d values 170-fold and 600-fold higher than for the *mta* promoter show an unexpected level of discrimination of MtaN for its own promoter. MtaN was

found to not specifically bind a modified PurR binding site (data not shown), thus despite significantly lower affinities for non-*mta* MerR-type promoters, they are still preferred over the *purF* operator.

Discussion

Much of the discrimination of MtaN for the *mta* promoter is likely due to DNA-binding elements in the winged-helix section of the protein. An alignment of this portion of MerR family members known to bind the DNA sites used in this study reveals a significant level of sequence homology in this region (Figure 5.4). Presumably, they share a similar DNA-binding mechanism. If this is the case, comparison of the DNA-binding regions of and binding sites for BmrR and MtaN may reveal possible modes of discrimination. The 3.0 Å resolution crystal structure of BmrR solved in complex with DNA and a drug was reported previously (Zheleznova-Heldwein *et al.* 2001). A 2.8 Å resolution crystal structure of a mutant BmrR (E253Q) has revealed a small number of previously unobserved contacts, including one hydrogen bond between the side chain of Lys 20, which is found on the recognition helix, and the O6 of guanine 5 (Dr. Joy Huffman, personal communication, unpublished results). The ability to visualize these contacts in the BmrR crystal structures is still limited by the statistical disorder in the half-site.

Most of the contacts that BmrR is observed to make to the backbone are likely to be conserved in the MtaN/*mta* complex. Because the peptide backbones of MtaN and BmrR overlay well in this region, contacts from the backbone amides of the BmrR residues Gly 9, Ala 21, Arg 43 and Leu 66 to the DNA backbone are likely to be

conserved by their MtaN counterparts (Lys 5, Thr 18, Arg 39 and Leu 62) (Figure 5.5). Hydrogen bonds by BmrR amino acid side chains to the DNA backbone also seem to be conserved, with one exception. The side chains of BmrR residues Tyr 25, Tyr 42, Arg 43 and Lys 60, which hydrogen bond to the DNA backbone, correspond to the side chains of MtaN residues Tyr 22, Tyr 38, Arg 39 and Lys 56, which are also able to contribute to hydrogen bonding. The only side-chain to backbone hydrogen bond observed in BmrR that would be impossible in an MtaN complex is that from Ser 41, which corresponds to a glycine in MtaN.

One of the base-specific contacts seen in the BmrR complex is an insertion of Tyr 42 into the minor groove of the DNA, where it makes van der Waals contacts with guanine 10 and a hydrogen bond to O2 of cytosine 10' (Zheleznova-Heldwein *et al.* 2001). This residue and the GC base pair are conserved in MtaN and both half-sites of the pseudo-palindromic *mta* operator, suggesting a conservation of these contacts. However, other BmrR-base contacts are not likely to be conserved between MtaN and BmrR. In the BmrR-*bmr* operator structure, van der Waals contacts are seen between Arg 23 and the C7 atom of thymine 7', and between Tyr 24 and adenine 2 and guanine 3 (Dr. Joy Huffman, personal communication, unpublished results). Both are major groove contacts from residues on the α2 helix, as is the hydrogen bond between the side chain of Lys 20 and the O6 of guanine 5. BmrR residues Lys 20, Arg 23 and Tyr 24 are not conserved in MtaN, where they correspond to Arg 17, His 20 and His21. Of the observed base contacts by these residues, only Arg 17 would be likely to make similar contacts in an MtaN/mta complex. Where BmrR Lys 20 forms a hydrogen bond to O6 of guanine 5,

the longer side chain of MtaN Arg 17 may be extended enough to contact the guanine 4, thymine 5 or guanine (or thymine) 6 of the *mta* operator to form hydrogen bonds with any one (or two) of these bases, all of which display hydrogen bond acceptors in the major groove. BmrR Arg 23 and Tyr 24 are longer than the corresponding His 20 and His 21 of MtaN, and from the BmrR structure, it does not appear that the histidine side chains would be able to make the same types of contacts observed for those side chains, however, they may be form hydrogen bonds to the DNA backbone or water-mediated bonds to the DNA-bases. Based upon these comparisons, if the MtaN/*mta* complex structure forms in a manner similar to the BmrR/*bmr* complex, the discrimination of MtaN is attributable to residues Arg 17, His 20 and His 21.

However, it is also possible that BmrR employs only one of several possible binding modes of these proteins. As discussed in the appendix, structures of MarA (Rhee *et al.* 1998) and Rob (Kwon *et al.* 2000) from *E. coli* bind DNA differently even though they are highly homologous in their DNA-binding regions (Figure A.3). If alternate binding modes are used by the MerR family, sequence comparisons of the residues involved in DNA-binding and the cognate binding sites will be misleading. However, even if alternate binding modes are used by the MerR family proteins, DNA-contacting residues are still likely to be contained in the winged-helix motif.

It is also possible that MtaN will be able to activate MerR promoters in *B. subtilis* not examined in this study. Only the resistance of *mta1* cells to drugs was tested in the original study (Baranova *et al.* 1999), however the promoters activated by MtaN may

include those involved in metal-resistance or other stress-response systems, thus activation by MtaN may create cells which are highly-resistant to a variety of stresses. Testing a Δ*merR*:*mta1* strain for its ability to survive mercury exposure, compared to a Δ*merR* strain, may reveal a higher level of mercury resistance. Alternatively, the mRNA of *mta1* cells could be probed for enrichment of message encoded by the genes of the other MerR family operons in *B. subtilis*. Because this mutation is not found to be naturally occurring, it is possible that it introduces a competitive disadvantage in cells under normal growth conditions.

The current study has shown that MtaN is a specific DNA-binding protein. The ability of MtaN to activate transcription of *bmr*, *blt* and *ydf* is probably due to high intracellular levels of the protein, rather than high-affinity binding to multiple promoter sequences. High levels of *mtaN* mRNA are seen in cells with the *mta1* mutation due to its autogenous regulation, presumably leading to high levels of MtaN. Once MtaN is produced, it specifically activates its own transcription, creating a positive feedback loop. Only when concentrations of MtaN have reached threshold levels will MtaN be able to compete effectively with BmrR and BltR for binding at the *bmr* and *blt* promoters. The same is presumably true at the *ydf* promoter with the hypothetical YdfL protein.

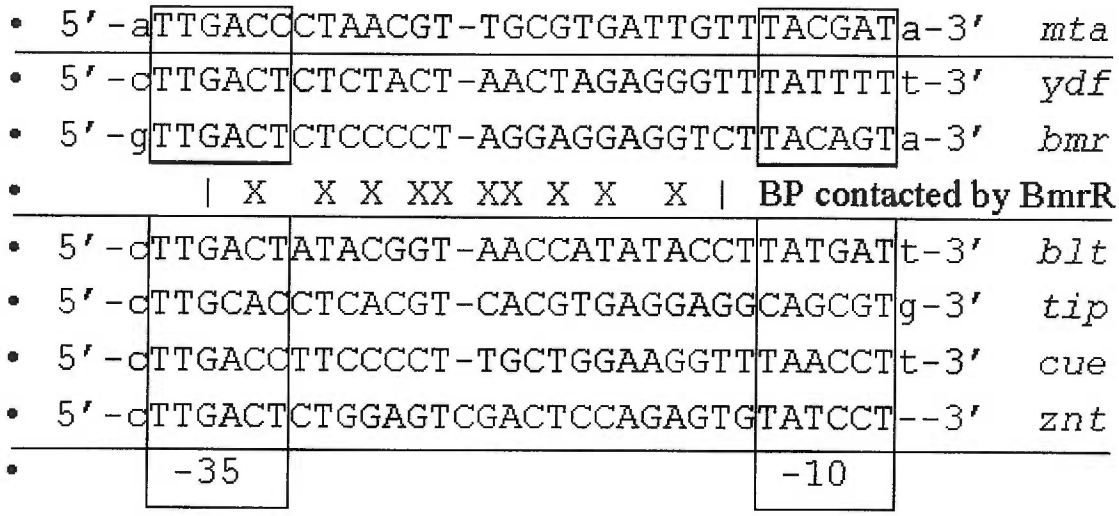

Table 5.1 Oligonucleotides used in FA binding studies. The strands shown were fluoresceinated at the 5' end and annealed to complementary non-flouresceinated strands as described in the materials and methods section. The sequences of each oligonucleotide were taken from the following references: *mta* (Baranova *et al.* 1999), *ydf* (Baranova *et al.* 1999), *bmr* (Baranova *et al.* 1999), *blt* (Baranova *et al.* 1999), *tip* (Holmes *et al.* 1993), *cue* (Outten *et al.* 2000), *znt* (Brocklehurst *et al.* 1999; Outten *et al.* 1999). The –10 and –35 promoter elements have been boxed and labelled.

Oligo name	K_d value, nM	Curve-fit error ± nM	Supplier
mta 33	4.6	0.8	O. E.
	8.1	2.6	Genosys
mta 30	6.1	1.7	Genosys
bmr 33	320	140	O. E.
	200	75	Genosys
blt 33	3,500	900	O. E.
	1,800	500	Genosys
ydf 33	240	40	O. E.
tip 33	1000	260	O. E.
cue 33	1000	160	O. E.
znt 33	3,500	1,100	O. E.

Table 5.2 Results from a representative set of MtaN binding experiments, as described in the text. This set of experiments was conducted on the same day with the same protein stock. (O. E. = Oligos Etc., Wilsonville, OR; Genosys = Sigma-Genosys, The Woodlands ,TX)

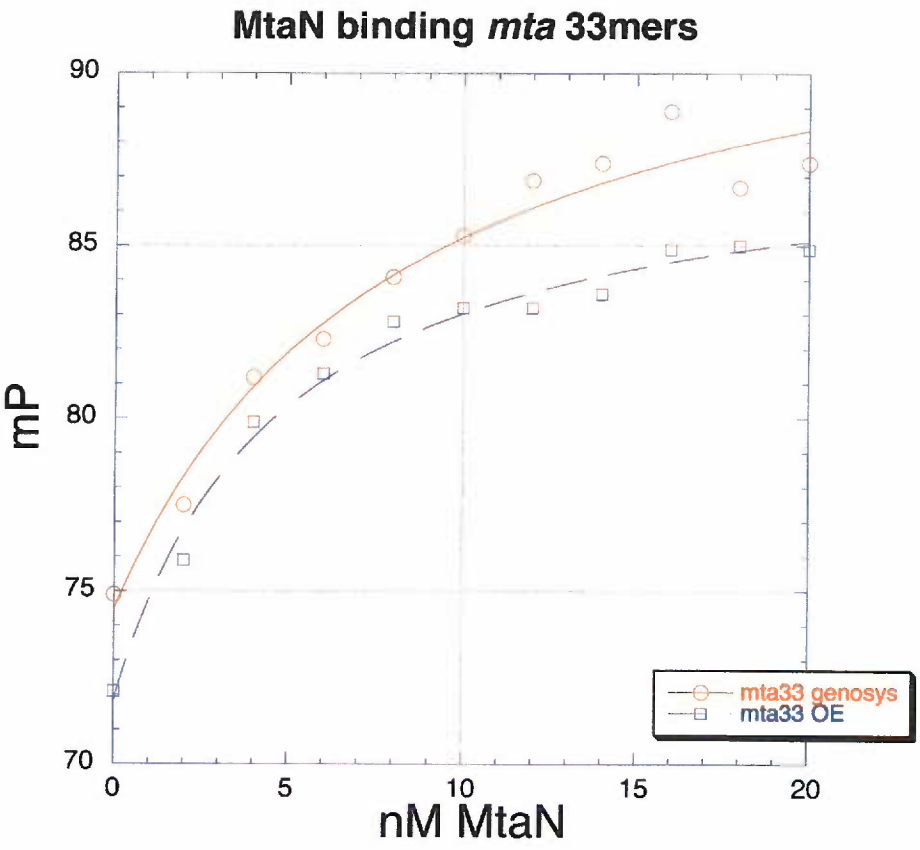

Figure 5.1 Binding of MtaN to the *mta* promoter. Two curves are shown for the oligonucleotides ordered from two different companies (OE = Oligos Etc., genosys = Sigma-Genosys). The K_d values derived from each curve are similar and given in the text and table 5.2.

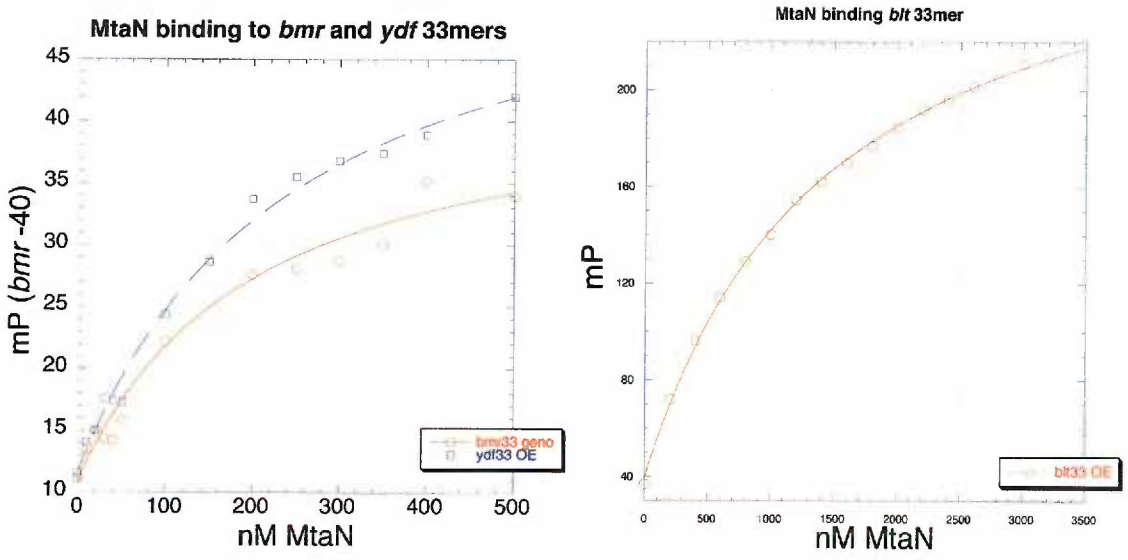

Figure 5.2 Binding of MtaN to operators of other *B. subtilis* MerR family members.

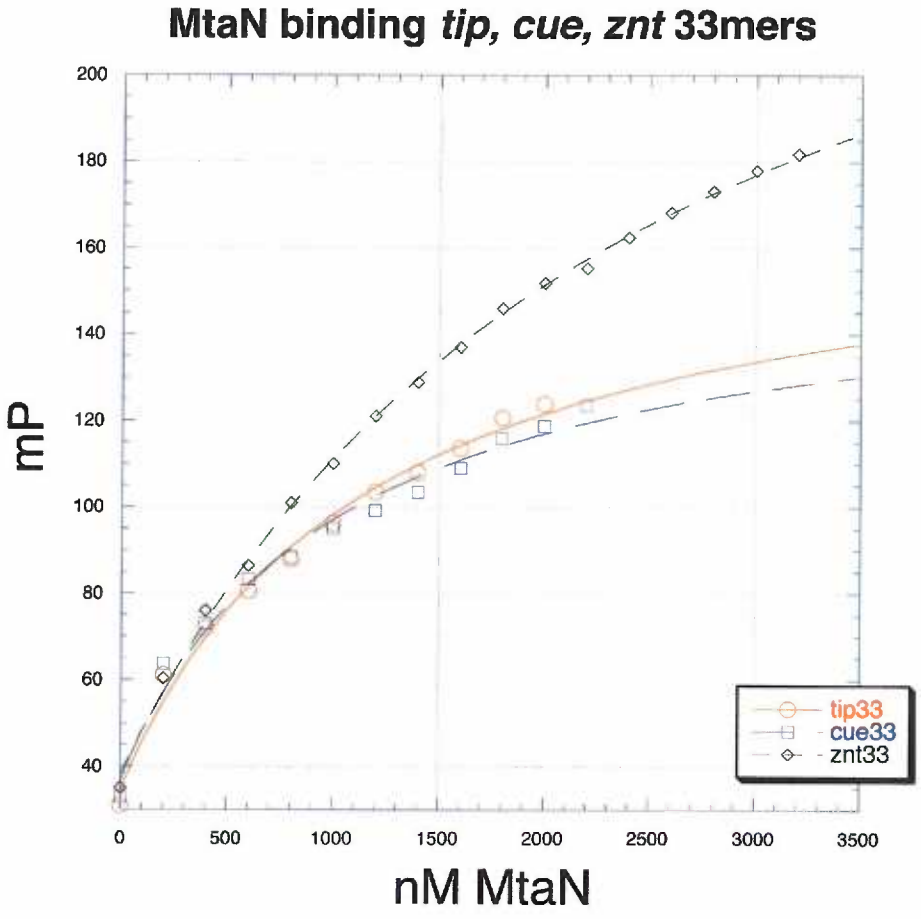

Figure 5.3 Binding of MtaN to MerR-type promoters from organisms other than *B. subtilis*.

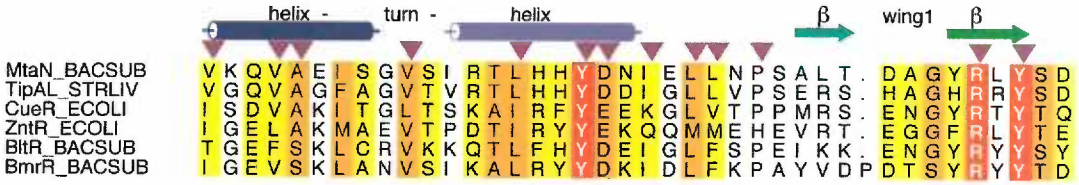

Figure 5.4 An alignment of the DNA-binding domains of selected MerR proteins known to bind the promoters used in this study. The secondary structure elements as found in MtaN are shown above the alignment. Purple arrowheads denote residues with side chains in the hydrophobic core of the domain. Conservation is color-coded from red (absolute) to white (less conserved).

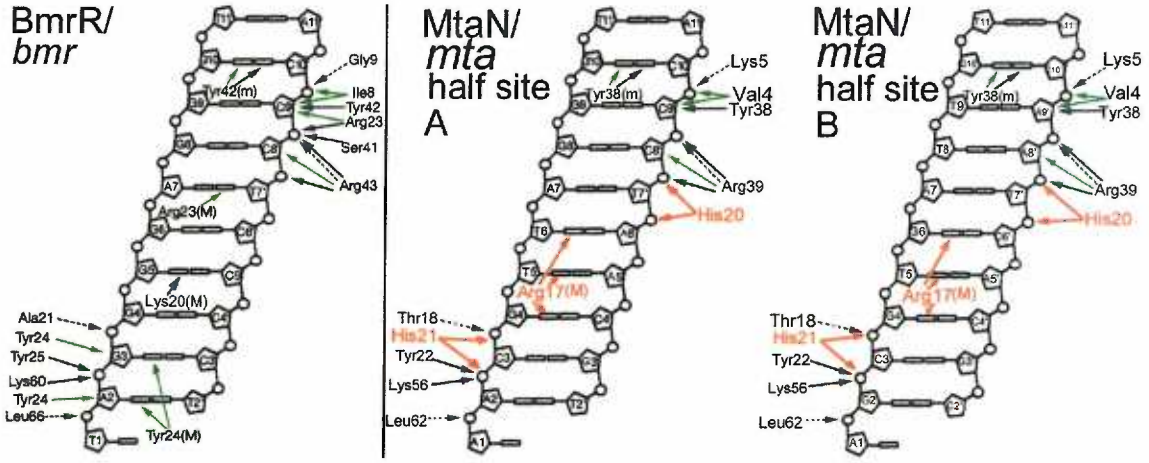

Figure 5.5 Schematic diagram of the observed BmrR-DNA contacts and hypothesized contacts between MtaN and the half-site of the *mta* promoter. The DNA is shown as a cylindrical projection, in which bases are depicted as rectangles, deoxyribose sugars are pentagons, and phosphates are circles. Hydrogen bonds are shown as blue arrows and van der Waals contacts as green arrows. Contacts likely to be conserved between MtaN and BmrR are shown as blue and green arrows, while predicted differential contacts are shown in red. Contacts in the major and minor grooves are labelled with "M" and "m", respectively. BmrR-*bmr* contact diagram is adapted from (Zheleznova-Heldwein *et al.* 2001).

Chapter 6

SUMMARY AND CONCLUSIONS

Bacterial multidrug resistance (MDR) is a large and growing human health concern. In *B. subtilis*, overexpression of Bmr and Blt multidrug transporters confers an MDR phenotype. Transcription of these transporters is specifically controlled by BmrR and BltR, respectively, both of which belong to the MerR family. Transcription of both is activated by a single mutant form of another MerR family member protein, MtaN. In order to understand further the mechanism and ability of MtaN to activate transcription at its own and three other promoters, and the general mechanism of activation by MerR family members, we carried out structural and biochemical studies on this protein.

Here, we report the structure of MtaN from *B. subtilis*, the first structure of a MerR family member protein to be solved in the absence of coactivator or DNA. BmrR, the specific regulator of the *bmr* gene, was solved by others in complex with DNA and a drug. Comparison of the structures of MtaN to BmrR reveals an overall structural conservation of the monomers. Specifically, a four-helix bundle is conserved as the DNA-contacting domain, and an antiparallel coiled-coil is conserved as a major dimerization interface. The hydrophobic core is also conserved in the DNA-contacting domain.

The overall conformations of the MtaN and BmrR dimers are shifted from each other. When monomers from each protein are overlaid, the dimer partners are in

significantly different positions. As compared to BmrR, the recognition helices of MtaN are further apart from each other, and the entire DNA-contacting domain of the dimer partner is rotated by approximately 15 degrees from the BmrR position. This correlates to the effective shortening and untwisting of the DNA observed in the BmrR/DNA/drug complex.

DNA-docking experiments reveal that the MtaN dimer is unable to bind the BmrR-activated DNA because its major groove recognition helices ($\alpha 2$) are too far apart and in the wrong orientation to fit into the *bmr* major grooves. MtaN is also unable to bind canonical B-form DNA because the $\alpha 2$ helices are tilted incorrectly to fit directly into adjacent major grooves and Wing 1 clashes with the DNA backbone. Thus at the least, MtaN requires minor structural adjustment in the twist of its $\alpha 2$ helices and more significant changes in the position of its β-sheet (Wing 1) in order to bind either DNA conformation.

These dissimilar dimeric conformations of the molecules have allowed us to propose a more detailed structural model of DNA-binding and activation by the MerR family of proteins. This proposal is predicated on the assumption that the BmrR-DNA complex represents the only activated form of the MerR family of proteins. In the first step of this proposal, MtaN binds to a B-like DNA conformation. We envision a model in which DNA-binding is likely to be concomitant with or followed by the breaking of the conserved Asp 23-Arg 39 salt bridge. The disruption would allow Arg 39 to contact the DNA phosphate backbone, possibly as observed in the BmrR-*bmr* promoter complex.

Additional structural changes would be transmitted through the coiled-coil and allow the MtaN conformation to maximize its DNA contacts. This in turn could elicit DNA conformational changes, which would result in an activated conformation of the MtaN-*mta* promoter complex that would closely resemble that of the BmrR-drug-*bmr* complex. Alternatively, the MtaN-*mta* complex may activate transcription differently than the BmrR-*bmr* complex, in which case some aspects of this model may be invalid.

The ability of MtaN to activate at least four promoters in *B. subtilis in vivo* led to the suggestion that Mta is a global regulator of multidrug resistance. To explore this possibility, we undertook binding studies of MtaN with a variety of promoters that are regulated by MerR family proteins. Previous work had suggested that the affinities of MtaN for the *bmr* and *blt* promoters would be within an order of magnitude of its affinity for its own promoter. We found this to not be the case. MtaN has a strong affinity for its own promoter and a significantly weaker affinity for other MerR-type promoters, including those of *bmr* and *blt*. We propose that MtaN is not a strong global activator. Instead, because MtaN is a constitutive activator and autogenously regulated, high levels of MtaN protein are produced through a positive feedback loop. These levels eventually become high enough to out-compete BmrR and BltR for binding to the *bmr* and *blt* promoters. Upon MtaN binding, the *bmr* and *blt* genes become constitutively active, resulting in constant production of the Bmr and Blt multidrug transporters and the MDR phenotype. Determination of the number of MtaN, BmrR and BltR molecules present in cells would allow us to more fully characterize this model.

In order to more fully understand the mechanism of transcriptional activation by MtaN, crystallization experiments are underway to determine the crystal structure of an MtaN/DNA complex. We have obtained data-quality crystals and begun solution of the structure. Preliminary results suggest that molecular replacement phasing may be sufficient to solve the structure of the complex. The complex crystals are grown with purified protein and 27 bp duplex DNA, encompassing the *mta* promoter, with single base overhangs. The molar ratio of protein dimer to DNA duplex is approximately 1:1. The crystals grow in space group C2 with a dimer in the asymmetric unit, suggesting that the statistical disorder seen in the BmrR complex will not be present in the MtaN complex. X-ray intensities data have been collected and processed to 2.7 Å resolution, nearly the same resolution used in the determination of the structure of the apo form. The solution of this structure will show whether MtaN utilizes the same mechanism of transcription activation as BmrR or a novel mode of binding and up-regulation.

In conclusion, the work set forth in this thesis has advanced our understanding of the mechanism of transcriptional activation by the MerR family of proteins, the regulation of MDR in *B. subtilis* and the conserved structure of the MerR family of proteins.

Appendix A

STRUCTURAL BIOLOGY OF MULTIDRUG EFFLUX AND MULTIDRUG RESISTANCE GENE REGULATORS

Michael H. Godsey, Ekaterina E. Zheleznova-Heldwein and Richard G. Brennan*

Department of Biochemistry and Molecular Biology
Oregon Health & Science University
Portland, OR 97201-3098
U.S.A.

*Corresponding author

Mailing address: Department of Biochemistry and Molecular Biology, L224
Oregon Health Sciences University
3181 S.W. Sam Jackson Park Road
Portland, Oregon 97201-3098, U.S.A.

Telephone: 503-494-4427; Fax: 503-494-8393; E-mail: brennanr@ohsu.edu

Running Title: Structural Biology of Multidrug Resistance

Key words: multidrug binding proteins/transporters/ABC-ATPases/bacterial transcriptional regulators/multidrug resistance

Introduction

Multidrug resistance (MDR) can be defined broadly as the ability of a cell to survive a seemingly lethal dose of more than one drug. Clearly, such resistance is a critical problem in the treatment of cancer and bacterial infection. Four general, but nonexclusive, mechanisms give rise to multidrug resistance: (1) detoxification by enzymatic modification or cleavage of drug, (2) genetic alteration of the intra or extracellular cellular target, (3) decreased permeability of the cell membrane and (4) active drug extrusion by multidrug transporters. Paramount to our understanding of MDR is the issue of recognition of structurally dissimilar substrates and how binding effects function. In bacteria many multidrug transporters are regulated directly (locally), by transcription factors, which also bind the substrates of these transporters, i.e., the drug can act as a transcriptional co-activator or inducer, or globally by activators that do not bind drugs (Grkovic *et al.* 2001a). The former class of proteins is of keen interest because they are more amenable to structural studies than the membrane-bound transporters and thus offer greater chance to obtain high resolution views of multidrug binding. The latter regulators are equally interesting as their DNA complexes reveal the mechanism of mdr transporter gene activation directly.

EmrE

The multidrug transporter EmrE is a member of the small multidrug resistance (SMR[1]) or MiniTexan family (Yerushalmi *et al.* 1995). The SMR pumps are the smallest MDR transporters; EmrE is 12 kDa per monomer. EmrE expels cationic lipophilic molecules, including ethidium, benzalkonium and tetraphenylphosphonium from the cell in a proton-dependent manner by utilization of the proton-motive force.

Thus, this protein as well as all other SMR members, functions as a H^+ antiporter (Yerushalmi *et al.* 1995). The oligomerization state of an active EmrE is consistent with a trimer and recent drug binding studies have shown 1 molecule of drug binds three EmrE monomers (Muth *et al.* 2000; Yerushalmi *et al.* 2000), although other oligomerization states could not be excluded.

Initial structural studies on EmrE by nuclear magnetic resonance (NMR) spectroscopy (Schwaiger *et al.* 1998), revealed this 110-residue protein to be comprised of four α helices (Schwaiger *et al.* 1998). These helices are tightly packed as established by cysteine-scanning mutagenesis in combination with N-ethylmaleimide labeling (Mordoch *et al.* 1999). In further support of a tightly packed structure, Fourier Transform Infrared (FTIR) spectroscopy data showed that the majority of the amide hydrogens of the polypeptide backbone do not readily exchange with solvent (Arkin *et al.* 1996). Taken together, these data imply that transmembrane helices of EmrE do not have any continuous aqueous cavities and are consistent with the binding data of other MD transporters, which show these pumps bind their substrates from the inner leaflet of the plasma membrane (Higgins *et al.* 1992; van Veen *et al.* 2000).

The NMR study supported earlier hydropathy plot analyses that EmrE has a single transmembrane acidic residue located in the middle of transmembrane segment 1 (Schuldiner *et al.* 1997). This membrane-embedded glutamate, Glu14, is the only charged residue within the protein that is important for activity (Yerushalmi *et al.* 2000). Extensive biochemical, kinetic, and mutational studies on EmrE have also revealed that Glu14 is the common site for substrate and proton binding, and that

protons compete directly with TPP^+ for binding (Muth *et al.* 2000; Yerushalmi *et al.* 2000). Thus, residue Glu14 is an integral part of the drug- and proton-binding site. The functional importance of a membrane-embedded carboxylate in other SMR family members is borne out by the finding that this glutamate is conserved among over 50 homologous transporters of this family (Schuldiner *et al.* 2001), e.g., the similarly located residue, Glu13, of SMR (QacC) from Staphylococcus aureus has been found to be essential for the multidrug transport activity of that multidrug pump (Grinius *et al.* 1994). The presence of a negative charge in the middle of a membrane-spanning domain provides the complement to the positively charged substrates of these proteins.

Our understanding of EmrE has been enhanced significantly by the recent determination of its projection structure to 7 Å resolution by two-dimensional cryo-electron microscopy (Tate *et al.* 2001). Although not at atomic resolution, the projection map revealed a surprise, EmrE, and likely all SMR proteins, is an asymmetric dimer and not trimeric as anticipated (Muth *et al.* 2000; Yerushalmi *et al.* 2000). The overall molecular dimensions of the dimer are 31 Å x 40 Å. Of particular interest is the disposition of the eight α helices of the dimer. Two are juxtaposed and lie perpendicular to the plane of the membrane. A second pair of helices is also adjacent but only one is located perpendicularly to the membrane whereas the other is tilted. These two sets of helices are separated by the four remaining α helices, which form an arc. These too are tilted with respect to the membrane. Such tilting of nearly half of the α helices of EmrE is consistent with FTIR data (Arkin *et al.* 1996).

The resolution of the current projection structure does not allow the specific assignment of helices to density. Clearly, 3-D structural information of higher resolution is needed to sort out the molecular details of multidrug transport of EmrE, foremost of which are the location and role of the Glu14 and how proton-substrate exchange affects the conformation of the transporter thus leading to expulsion of drug from the cell.

P-glycoprotein and ABC transporters

The first discovered and most studied multidrug-binding protein is the mammalian P-glycoprotein, an ATP-binding cassette (ABC) transporter, which is found to expel anti-cancer drugs. ABC transporters comprise a superfamily of ATP-dependent membrane pumps. Studies of *E. coli* suggest that 5% of total its genome codes for ABC transporters (Linton *et al.* 1998). A typical ABC transporter consists of two membrane-spanning domains, composed of six transmembrane (TM) segments each, and two ATP or Nucleotide Binding Domains (NBD). These domains can originate from separate proteins or be part of a single polypeptide chain.

Limited structural information is available for any intact multidrug transporter in the ABC family. An electron microscopy study determined the overall structure of P-glycoprotein to 25 Å resolution (Rosenberg *et al.* 1997). The authors describe the shape as toroidal with 6-fold symmetry and a diameter of 10 nm. A large central pore is seen that is open to the lipid phase, but closed to the cytoplasm. Such an architecture would be consistent with the current thought that most drugs bind multidrug transporters in the inner leaflet of the lipid bilayer rather than the more polar cytoplasm. Unfortunately, the resolution is insufficient to determine a mode or site of multidrug binding, however

the nucleotide binding domains are likely to be the two lobes, which are exposed to the cytoplasm. A higher resolution model of a homologous protein, *E. coli* MsbA, gives significantly more information. This model was discussed in chapter one of this thesis.

TolC

TolC is a bacterial outer-membrane protein that is part of the type I efflux machinery, which allows for the direct transport of proteins and toxic compounds from the cytosol and inner membrane to the extracellular space. The absence of periplasmic intermediates in the TolC-mediated removal of substrates suggests transport across two membranes is accomplished in one step. TolC is not a multidrug-binding protein per se, but critical for the drug-efflux machinery since it serves as the outflow tube of multidrug pumps. Electron microscopy studies found that TolC is trimeric with an outside diameter of 58 Å and large central hole (Koronakis *et al.* 1997). Recently the crystal structure of this protein at 2.5 Å resolution has become available (Koronakis *et al.* 2000). The structure reveals a tube, which is long enough to bridge the periplasm from the inner membrane, where it interacts with RND family member pumps like MexX/Y (Mine *et al.* 1999) or AcrA/B (Nikaido 2001; Nikaido *et al.* 2001), to the outer membrane where TolC carries its own outer membrane anchor in the form of a β-barrel.

The structure of TolC reveals a previously unseen architecture with a helical tube, ~100 Å long, comprised of 12 continuous and pseudo-continuous α-helices topped by a 12-stranded β-barrel (Figure 1). The 12 elements of secondary structure in the homotrimer originate equally from the three molecules. Extra secondary structure appears around the outside of the α-helical tube, somewhat like a metal strap around a

barrel. The inside of the trimer is open to the outside of the cell, with an internal diameter of ~35 Å for most of its length. The center is mostly hydrophobic near the distal (extracellular) end, with the proximal end being significantly electropositive. This gradient might serve to attract neutral or electronegative toxins from the inner-membrane pump, and then allow these amphipathic molecules to move along the length of the pipe as solvent is able to shield the electrostatic forces that would have pulled the molecule into TolC.

The role of TolC in multidrug extrusion appears to be passive offering a general exit route, in which both large and small molecules are able to avoid the periplasmic space. Structures of TolC bound to a variety of small molecule ligands will be needed to reveal the principles of multidrug recognition by this macromolecular chunnel.

BmrR / BRC / MtaN

The first view of a multidrug-binding site was revealed in crystallographic studies on a cytosolic protein that, like the multidrug transporters, is capable of recognizing multiple ligands with dissimilar structures. This protein is BmrR (Bacterial multidrug resistance Regulator), a MerR family member (Summers 1992) from the bacterium Bacillus subtilis. Like other MerR family members, BmrR contains 3 domains: an N-terminal DNA-binding domain, a linker/dimerization region, and a C-terminal co-activator (drug)-binding domain. Upon binding one of its structurally unrelated lipophilic cationic ligands, BmrR activates transcription of the multidrug transporter gene, bmr (Ahmed *et al.* 1994). Many of these ligands (coactivators) are also substrates of the multidrug transporter Bmr.

To facilitate structural studies, a 159-residue ligand-binding domain, designated BRC (for BmrR C-terminus), which binds drugs with the same affinities as the full-length BmrR (Markham *et al.* 1996), was crystallized (Zheleznova *et al.* 1997) and solved in its drug-free and tetraphenylphosphonium (TPP^+) bound forms were determined (Zheleznova *et al.* 1999). The BRC TPP^+ structure unveiled an internal multidrug-binding pocket that is lined with hydrophobic and aromatic amino acids, which participate in van der Waals and stacking interactions with four phenyl rings of the TPP^+ molecule. Importantly, the bottom of the pocket features a glutamate residue, Glu253, buried in the hydrophobic core of the protein. The positively charged TPP^+ makes an electrostatic interaction with the negatively charged carboxylate group of Glu253, an interaction that is enhanced by the low dielectric constant of the protein interior (Zheleznova *et al.* 2000). In the drug-free, or apo form, this binding pocket is completely shielded from the surrounding medium by a short α-helix. In fact, the binding pocket cannot be detected in the apo structure as Glu253 is completely buried within the solvent inaccessible, hydrophobic core; where its negatively charged carboxylate group is "neutralized" by hydrogen bonds to the hydroxyl groups of three tyrosine residues. To allow ligand access this binding site, the helical shield undergoes a helix-to-coil transition and moves away from the protein. The unwound helix is disordered in crystals of the BRC-TPP^+ complex and is highly susceptible to trypsinolysis in solution, thus indicating its conformational flexibility. The interactions between BRC and TPP^+ suggest a similar binding mode for other hydrophobic cationic ligands of BRC/BmrR. Since all BmrR ligands are hydrophobic cations, the

electrostatic interaction between positively charged ligands and the negatively charged glutamate is key to the cation selectivity of BmrR. Indeed, the conservative replacement of Glu253 with the isosteric but electroneutral amino acid glutamine abolishes drug binding but does not perturb the global structural integrity of the protein (Zheleznova *et al.* 1999) (Huffman, Zheleznova and Brennan, unpublished data).

The ligand selectivity of BRC depends not only on the presence of the negatively charged glutamate but also on the architecture and chemistry of the binding site. The hydrophobic and aromatic amino acids in the binding site are arranged in a way that precludes the binding of most positively charged molecules, for instance, no carbonyl oxygens are available to replace the hydration shell of a divalent cation (Zheleznova *et al.* 1999). Although the degree of structural flexibility of the binding site is unknown, the hydrophobic side chains likely rotate to accommodate drugs with different geometry and to bind maximize their affinity. Furthermore, mutational analysis of the BRC residues directly participating in binding of ligands showed that each mutation affected the binding affinity of different ligands in a different way (Vazquez-Laslop *et al.* 1999). This suggests that although ligands bind in the same location within the binding pocket, each ligand forms a distinct set of hydrophobic and van der Waals contacts with the residues in the binding site, but the binding pocket does not undergo dramatic rearrangement. The structures of additional BRC-drug complexes should clarify this issue.

Recently, the structures of full-length BmrR in complex with DNA and a drug (Heldwein *et al.* 2001) and MtaN (multidrug transporter activation, N-terminus) also from B. subtilis, have been solved (Godsey *et al.* 2001) (Figure 2). BmrR was the first

intact MerR family member to have its structure determined, and the structure of MtaN, a constitutive activator of transcription, was the first of a MerR family member solved in the absence of DNA or coactivator. Drug binding by BmrR is very similar to that of the BRC with the exception of the possible, and puzzling, involvement of residue Asp47, which is located in the DNA binding domain. Site directed mutagenesis of Asp47 and drug binding studies should elucidate its role.

Comparison of BmrR and MtaN reveals a similar winged helix DNA-binding motif, part of which is contained within a structurally conserved four-helix bundle. However, a crucial difference between the two proteins is evident in their dimer conformations. The major-groove binding helices (the "recognition" helices) are separated by 33.3 Å in the MtaN structure but by only 30.6 Å in the DNA- and coactivator-bound form of BmrR, reflecting shortening of the effective length of the consecutive major grooves of BmrR-bound bmr promoter. In addition, there is a 15° twist of the dimer partners, which is mediated through the 8 (MtaN) or 11-turn (BmrR) dimerization helices, an antiparallel coiled coil (Godsey *et al.* 2001), that causes a 7.5 Å shift in the relative positions of the major-groove binding helices. These differences support a mechanism of DNA activation by MerR family members in which DNA- and coactivator-binding cause a shift in the dimer conformation of the protein that leads to disruption of the central A-T base pair and shortening and under-twisting of the DNA (Summers 1992; Ansari *et al.* 1995; Outten *et al.* 1999). The resultant DNA structure changes the orientation of the –35 and –10 promoter elements to allow the productive binding of RNA polymerase and open complex formation. The structure of an MtaN-

mta promoter complex is expected to have a similar DNA distortion but will also delineate any protein-specific differences.

MarA/Rob/MarR

MarA is a member of the AraC family of transcriptional regulators (Martin *et al.* 2001) that activates over a dozen genes comprising the E. coli mar (multiple antibiotic resistance) regulon (Alekshun *et al.* 1997; Barbosa *et al.* 2000). Interestingly, MarA is a constitutive activator in that this protein does not bind antibiotics as part of its activation mechanism. MarA is a 129 amino acyl residue monomer, and binds asymmetric 20 base-pair operators (Martin *et al.* 1996), which is in contrast to most prokaryotic transcriptional regulators, including BmrR and MtaN, that are dimers and bind pseudo-palindromic DNA (Harrison *et al.* 1990; Steitz 1990).

The structure of MarA, the first of its family, bound to the marA operator revealed the basis of its monomeric DNA-binding properties (Rhee *et al.* 1998). The protein consists of seven α helices, six of which comprise two three-helix bundles each containing an HTH motif. The remaining helix connects the two HTH motifs thus creating a monomer with two DNA-reading heads. Each independent HTH binds adjacent major grooves and the sequentially different recognition helices make a distinct set of contacts to the DNA bases, thus explaining the ability of the MarA monomer to bind asymmetric operator sequences. To gain such discrimination otherwise would require heterodimerization such as that displayed by the eukaryotic transcriptional regulators Myc-Max (Cole 1991; Brownlie *et al.* 1997), Fos-Jun (Glover *et al.* 1995) and AHR-ARNT (Hankinson 1995).

In order to bind a cognate DNA site, MarA must bend the DNA significantly. Bending occurs because the recognition helices of MarA are separated by only 27 Å, and conformational flexibility is limited by the linker helix, which constrains the distance between the two HTH motifs. Thus, to accommodate the shorter distance and bind consecutive major grooves, MarA "pulls" the major grooves towards itself, causing kinks in the DNA at each HTH by narrowing the minor groove. These localized dual kinks result in a global DNA bend of ~35° (Rhee *et al.* 1998).

Recently, a second AraC family member, the E. coli Rob, has been solved bound to DNA (Kwon *et al.* 2000). Rob, like MarA, can activate the mar operon when overexpressed (Ariza *et al.* 1995). However, unlike MarA, Rob contains an additional ~200 amino acyl residue C-terminal domain of unknown function but which is structurally similar to the E. coli galactose-1-phosphate uridyltransferase (Thoden *et al.* 1997). As expected, MarA and the N-terminal, DNA-binding domain of Rob, the sequences of which are 51% identical, are similar structurally. Indeed, an overlay of all the main-chain atoms of their conserved domains results in an r.m.s. deviation of 0.9 Å (Kwon *et al.* 2000). Yet, their modes of DNA-binding are quite different. Only the N-terminal HTH motif of Rob is inserted into the major groove. The C-terminal HTH sits on the surface of the double helix and contacts the DNA backbone. This conformation allows Rob to bind unbent B-form DNA, as is seen in the crystal structure. However, Rob is also known to bind bent DNA (Jair *et al.* 1996). The mechanistic implications of this altered DNA-binding mode are unclear and will require additional Rob-DNA structures.

The common feature of DNA recognition by MarA and Rob is the insertion of helix 2 of the N-terminal HTH motifs and, in MarA the C-terminal HTH, into the major groove of the DNA, whereby these helices make similar but limited specific contacts to the bases. In the three helices, a conserved Arg (MarA: residues 46, 96; Rob: residue 40) makes a hydrogen bond to a guanine base. In MarA these residues also make hydrogen bonds to other bases as well. In the C-terminal HTH of Rob, the homologous residue, Arg90, interacts with the DNA phosphate backbone. The only other specific hydrogen bonds seen in these domains are two water-mediated bonds from MarA residue Thr93 to a thymine and a cytosine. The other HTH-DNA contacts are comprised of van der Waals interactions and nonspecific backbone contacts.

Whereas MarA, and perhaps Rob, responds to multidrug intrusion, this constitutive activator does not bind drugs. Rather, its expression is negatively regulated by MarR (Alekshun *et al.* 1997), a repressor that acts on the marRAB locus. The structure of MarR was the first of the MarR family to be determined (Alekshun *et al.* 2001). MarR is a 144-residue α/β protein composed of 6 α-helices and 3 β–strands. The dimer interface is formed mainly by the N- and C- termini of the monomers, which bury 3570 $Å^2$ of accessible surface area. The DNA-binding regions are found in the middle of the protein sequence, and MarR uses a winged-helix to bind DNA. Overall, the dimer appears to have the shape of a highway-warning triangle (Fig 3). Interestingly, though the related EmrR binds one drug molecule per dimer (Brooun *et al.* 1999), the structure of MarR was solved with two salicylate molecules bound per monomer. Thus, although salicylate is known to inhibit MarR (Alekshun *et al.* 1999), it

remains unclear if either one or both molecules is binding a biologically significant drug-binding site. A ligand-free form, as well as forms bound to DNA and bound to another known co-repressor, are necessary to more fully understand the action of this repressor.

Perspective

Clearly we are at the early stages of elucidating the structural underpinnings of multidrug recognition, how these toxic molecules regulate gene expression and the mechanisms by which cells remove them. Major structural questions remain. For instance, are multidrug binding pockets limited in their design or have multiple scaffolds arisen? How similar are the drug binding mechanisms of cytosolic and membrane-bound multidrug binding proteins? How does ATP hydrolysis or the proton motive force effect drug extrusion? Over the next few years the structures of additional drug-bound mdr regulators should address these questions in part and as importantly delineate their mechanisms in mdr gene regulation. Of course, the cardinal aim of structural biologists will be the determination of the high resolution structure of a multidrug transporter. To this end, low resolution structures of P-glycoprotein (Rosenberg *et al.* 1997) and MRP1 (Rosenberg *et al.* 2001), both of which are members of the ATP-binding cassette transporter family, have been studied by electron microscopy. Each reveals a similar key feature, a pore that is opened at the outer membrane but closed off at the inner membrane, presumably by their nucleotide binding domains. Interestingly, while P-glycoprotein is a monomer, MRP1 appears dimeric, the biological significance of which is unclear. More recent, a 6 Å resolution projection structure of the oxalate transporter (OxlT), a member of the major facilitator

superfamily, was described and revealed an opening, its substrate-transport channel, centered about a pseudo two-fold that relates its two sets of six transmembrane helices (Heymann *et al.* 2001). Although OxlT is not a multidrug transporter, many MDR transporters belong to this family and their structures are expected to share many general features. However, because multidrug transporters extract their substrates from the inner leaflet, drug entrance to the efflux channel will likely require some, perhaps subtle, structural differences. Thus, although the structure of P-glycoprotein might be the "holy grail", many structures of multiple transporters from diverse families complemented by biochemical and biophysical studies will be needed to understand the biology of these fascinating and medically important molecules.

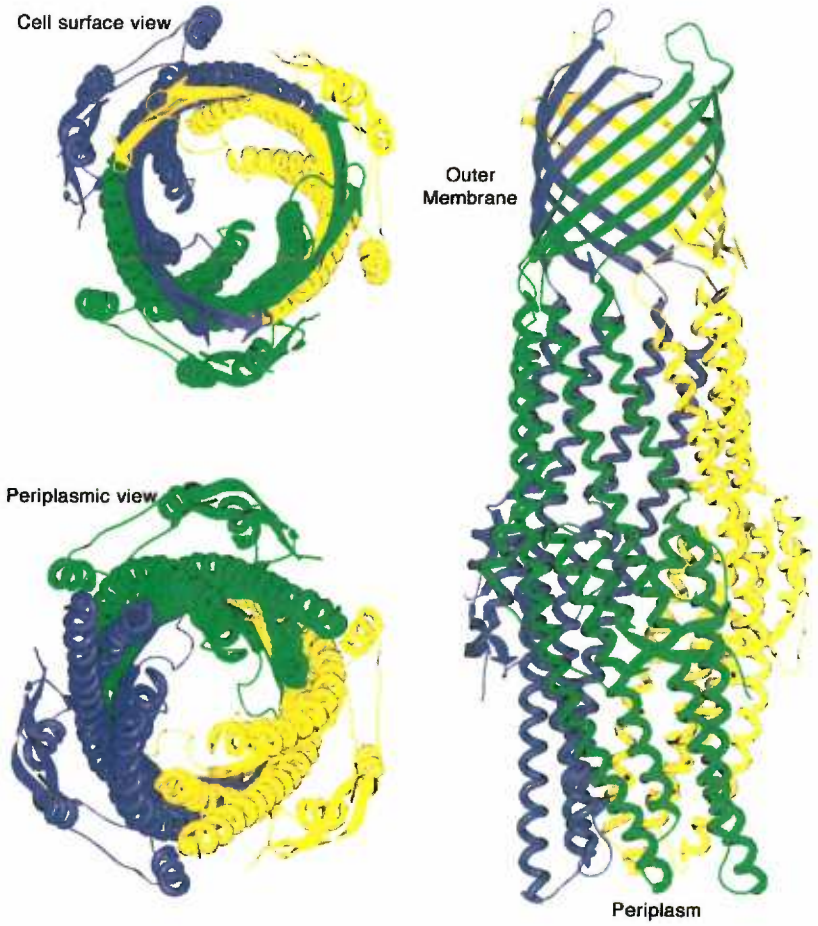

Figure A.1. The TolC "chunnel" protein model, as seen from the side and the ends. The β-barrel is in the outer membrane and the 12-membered helical tube is inserted into the periplasm. Monomers are colored green, blue and yellow to show the orientation of the monomers in the trimer structure.

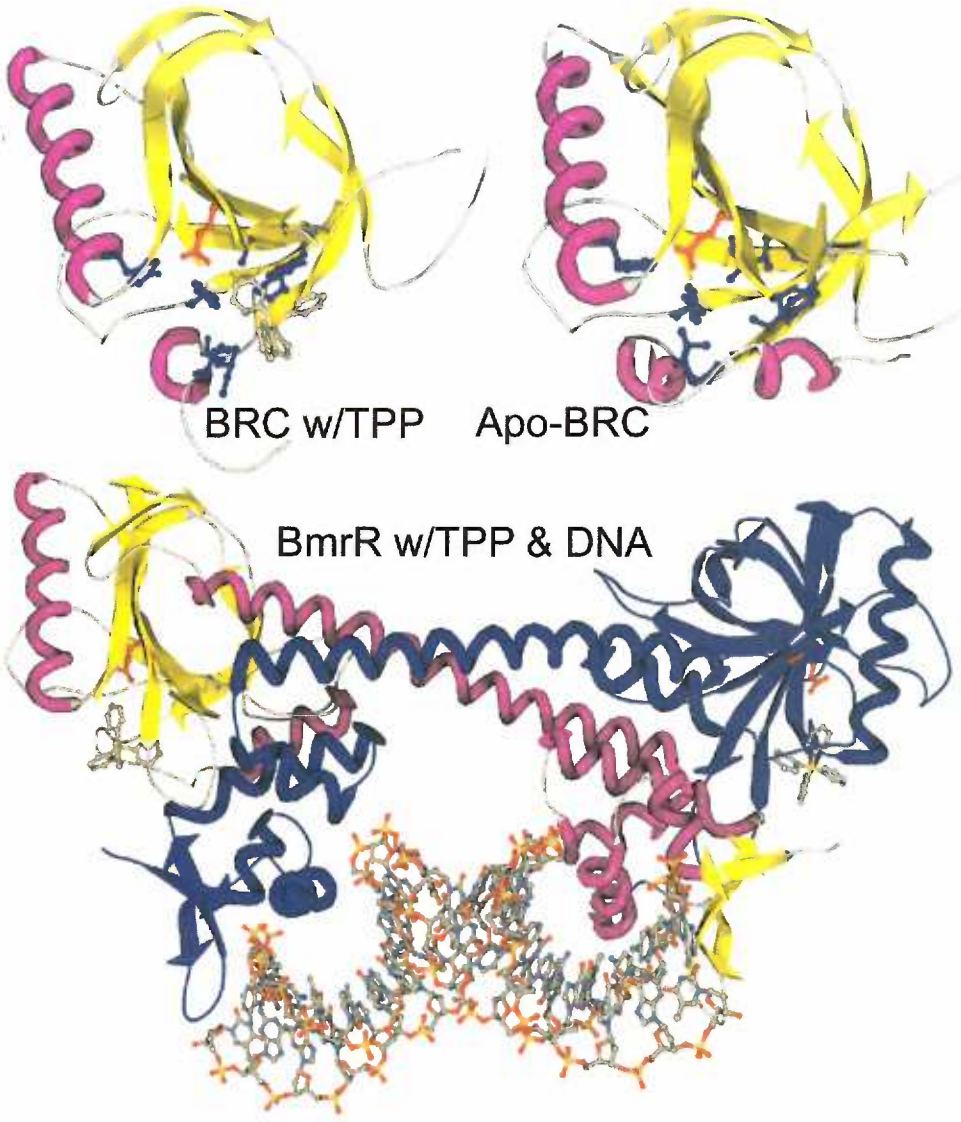

Figure A.2. Crystal structure models of BRC with and without bound tetraphenylphosphonium (TPP^+) and full-length BmrR bound to TPP^+ and DNA. TPP^+ is bound similarly in both BRC and BmrR. A buried glutamate, which neutralizes the positive charge on TPP^+, is shown in red and hydrophobic residues lining the drug-binding pocket are shown in blue in the BRC models.

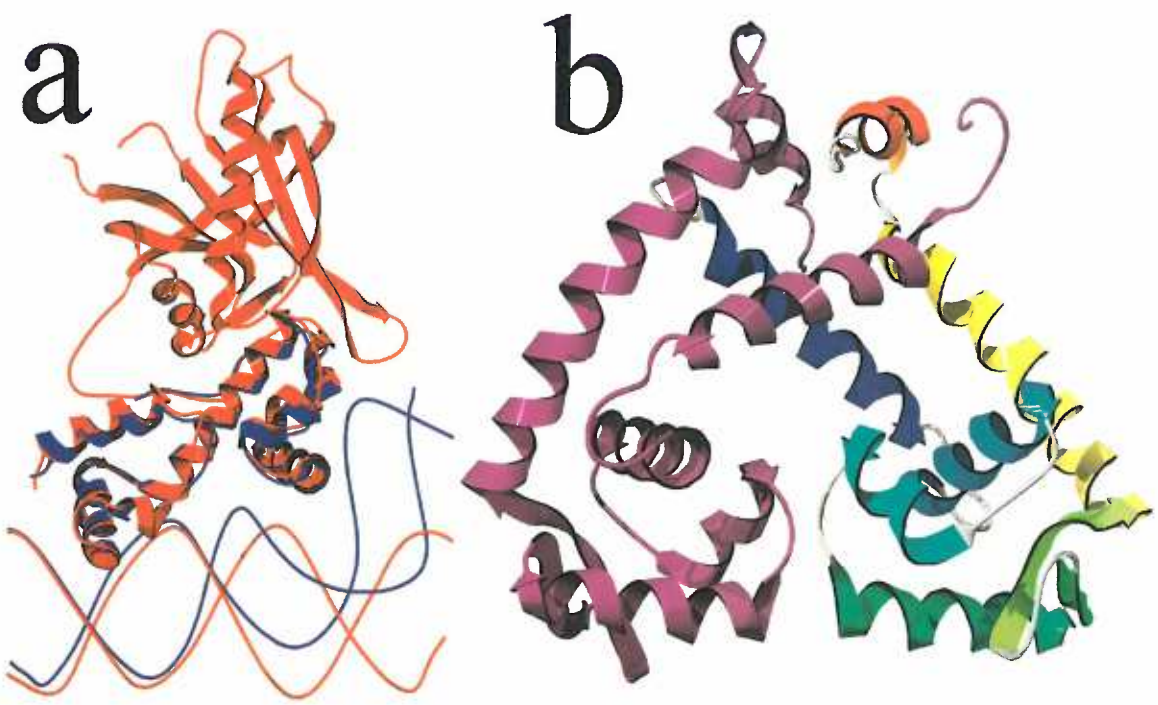

Figure A.3. (a) Aligned models of MarA (blue) and Rob (red) bound to their respective DNA sites. The MarA model shows DNA bent towards the protein such that both HTH motif are inserted into consecutive major grooves. In contrast, the Rob model shows one HTH binding a major groove, while the other sits on the phosphate backbone of the unbent DNA. (b) Dimer model of MarR. One monomer is colored from blue (N-terminal) to red (C-terminal). One recognition helix is green and located at the bottom of the monomer in this view.

Literature Cited:

Abrahams, J. P. (1997). Bias Reduction in Phase Refinement by Modified Interference Functions: Introducing the gamma Correction. Acta Crystallogr D **53**(4): 371-376.

Ahmed, M., Borsch, C. M., Taylor, S. S., Vazquez-Laslop, N. and Neyfakh, A. A. (1994). A protein that activates expression of a multidrug efflux transporter upon binding the transporter substrates. J Biol Chem **269**(45): 28506-13.

Ahmed, M., Lyass, L., Markham, P. N., Taylor, S. S., Vazquez-Laslop, N. and Neyfakh, A. A. (1995). Two highly similar multidrug transporters of Bacillus subtilis whose expression is differentially regulated. J Bacteriol **177**(14): 3904-10.

Alekshun, M. N. and Levy, S. B. (1997). Regulation of chromosomally mediated multiple antibiotic resistance: the mar regulon. Antimicrob Agents Chemother **41**(10): 2067-75.

Alekshun, M. N. and Levy, S. B. (1999). The mar regulon: multiple resistance to antibiotics and other toxic chemicals. Trends Microbiol **7**(10): 410-3.

Alekshun, M. N., Levy, S. B., Mealy, T. R., Seaton, B. A. and Head, J. F. (2001). The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution. Nat Struct Biol **8**(8): 710-714.

Alt, F. W., Kellems, R. E., Bertino, J. R. and Schimke, R. T. (1992). Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. 1978. Biotechnology **24**: 397-410.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol **215**(3): 403-10.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**(17): 3389-402.

Amyes, S. G., Towner, K. J., Carter, G. I., Thomson, C. J. and Young, H. K. (1989). The type VII dihydrofolate reductase: a novel plasmid-encoded trimethoprim-resistant enzyme from gram-negative bacteria isolated in Britain. J Antimicrob Chemother **24**(2): 111-9.

Anderson, J. E., Ptashne, M. and Harrison, S. C. (1987). Structure of the repressor-operator complex of bacteriophage 434. Nature **326**(6116): 846-52.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y. and Matthews, B. W. (1981). Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. Nature **290**(5809): 754-8.

Ansari, A. Z., Bradner, J. E. and O'Halloran, T. V. (1995). DNA-bend modulation in a repressor-to-activator switching mechanism. Nature **374**(6520): 371-5.

Ansari, A. Z., Chael, M. L. and O'Halloran, T. V. (1992). Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. Nature **355**(6355): 87-9.

Aramaki, H., Yagi, N. and Suzuki, M. (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.

Ariza, R. R., Li, Z., Ringstad, N. and Demple, B. (1995). Activation of multiple antibiotic resistance and binding of stress- inducible promoters by Escherichia coli Rob protein. J Bacteriol **177**(7): 1655-61.

Arkin, I. T., Russ, W. P., Lebendiker, M. and Schuldiner, S. (1996). Determining the secondary structure and orientation of EmrE, a multi- drug transporter, indicates a transmembrane four-helix bundle. Biochemistry **35**(22): 7233-8.

Baranova, N. N., Danchin, A. and Neyfakh, A. A. (1999). Mta, a global MerR-type regulator of the Bacillus subtilis multidrug- efflux transporters. Mol Microbiol **31**(5): 1549-59.

Barbosa, T. M. and Levy, S. B. (2000). Differential expression of over 60 chromosomal genes in Escherichia coli by constitutive expression of MarA. J Bacteriol **182**(12): 3467-74.

Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng **6**(1): 37-40.

Bewley, C. A., Gronenborn, A. M. and Clore, G. M. (1998). Minor groove-binding architectural proteins: structure, function, and DNA recognition. Annu Rev Biophys Biomol Struct **27**: 105-31.

Blow, D. M. and Crick, F. H. C. (1959). The treatment of errors in the isomorphous replacment method. Acta Crystallogr **12**: 794-802.

Bolhuis, H., van Veen, H. W., Molenaar, D., Poolman, B., Driessen, A. J. and Konings, W. N. (1996). Multidrug resistance in Lactococcus lactis: evidence for ATP-dependent drug extrusion from the inner leaflet of the cytoplasmic membrane. Embo J **15**(16): 4239-45.

Bolhuis, H., van Veen, H. W., Poolman, B., Driessen, A. J. and Konings, W. N. (1997). Mechanisms of multidrug transporters. FEMS Microbiol Rev **21**(1): 55-84.

Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. Cell **74**(5): 773-6.

Brennan, R. G. and Matthews, B. W. (1989). The helix-turn-helix DNA binding motif. J Biol Chem **264**(4): 1903-6.

Brennan, R. G., Roderick, S. L., Takeda, Y. and Matthews, B. W. (1990). Protein-DNA conformational changes in the crystal structure of a lambda Cro-operator complex. Proc Natl Acad Sci U S A **87**(20): 8165-9.

Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. Acta Crystallogr A **32**: 832-847.

Brocklehurst, K. R., Hobman, J. L., Lawley, B., Blank, L., Marshall, S. J., Brown, N. L. and Morby, A. P. (1999). ZntR is a Zn(II)-responsive MerR-like transcriptional regulator of zntA in Escherichia coli. Mol Microbiol **31**(3): 893-902.

Brooun, A., Tomashek, J. J. and Lewis, K. (1999). Purification and ligand binding of EmrR, a regulator of a multidrug transporter. J Bacteriol **181**(16): 5131-3.

Brownlie, P., Ceska, T., Lamers, M., Romier, C., Stier, G., Teo, H. and Suck, D. (1997). The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. Structure **5**(4): 509-20.

Brünger, A. T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature **355**: 472-474.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998). Crystallography &

NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-21.

Brünger, A. T., Krukowski, A. and Erickson, J. W. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. Acta Crystallogr A **46**(7): 585-593.

Busch, S. J. and Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. Trends Genet **6**(2): 36-40.

Caguiat, J. J., Watson, A. L. and Summers, A. O. (1999). Cd(II)-responsive and constitutive mutants implicate a novel domain in MerR. J Bacteriol **181**(11): 3462-71.

Cantor, C. A. and Schimmel, P. R. (1980). Biophysical chemistry: part II: techniques for the study of biological structure and function. New York, W. H. Freeman & Company.

Chang, G. and Roth, C. B. (2001). Structure of MsbA from E. coli: A Homolog of the Multidrug Resistance ATP Binding Cassette (ABC) Transporters. Science **293**(5536): 1793-800.

Chiu, M. L., Folcher, M., Griffin, P., Holt, T., Klatt, T. and Thompson, C. J. (1996). Characterization of the covalent binding of thiostrepton to a thiostrepton-induced protein from Streptomyces lividans. Biochemistry **35**(7): 2332-41.

Clark, K. L., Halay, E. D., Lai, E. and Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature **364**(6436): 412-20.

Cole, M. D. (1991). Myc meets its Max. Cell **65**(5): 715-6.

Cole, S. P., Bhardwaj, G., Gerlach, J. H., Mackie, J. E., Grant, C. E., Almquist, K. C., Stewart, A. J., Kurz, E. U., Duncan, A. M. and Deeley, R. G. (1992). Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. Science **258**(5088): 1650-4.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D **50**: 760-763.

Comess, K. M., Shewchuk, L. M., Ivanetich, K. and Walsh, C. T. (1994). Construction of a synthetic gene for the metalloregulatory protein MerR and analysis of regionally mutated proteins for transcriptional regulation. Biochemistry **33**(14): 4175-86.

Conte, L. L., Chothia, C. and Janin, J. (1999). The atomic structure of protein-protein recognition sites. J Mol Biol **285**(5): 2177-98.

Davies, D. R. and Cohen, G. H. (1996). Interactions of protein antigens with antibodies. Proc Natl Acad Sci U S A **93**(1): 7-12.

Davies, J. (1994). Inactivation of antibiotics and the dissemination of resistance genes. Science **264**(5157): 375-82.

Diederichs, K., Diez, J., Greller, G., Muller, C., Breed, J., Schnell, C., Vonrhein, C., Boos, W. and Welte, W. (2000). Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon thermococcus litoralis. Embo J **19**(22): 5951-61.

Dodd, I. B. and Egan, J. B. (1987). Systematic method for the detection of potential lambda Cro-like DNA- binding regions in proteins. J Mol Biol **194**(3): 557-64.

Dodd, I. B. and Egan, J. B. (1990). Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res **18**(17): 5019-26.

Drenth, J. A. (1999). Principles of protein X-ray crystallography. New York, Springer-Verlag.

Fath, M. J. and Kolter, R. (1993). ABC transporters: bacterial exporters. Microbiol Rev **57**(4): 995-1017.

Filgueiras, M. H. and Op den Kamp, J. A. (1980). Cardiolipin, a major phospholipid of Gram-positive bacteria that is not readily extractable. Biochim Biophys Acta **620**(2): 332-7.

Frantz, B. and O'Halloran, T. V. (1990). DNA distortion accompanies transcriptional activation by the metal- responsive gene-regulatory protein MerR. Biochemistry **29**(20): 4747-51.

Fujinaga, M., Berthet-Colominas, C., Yaremchuk, A. D., Tukalo, M. A. and Cusack, S. (1993). Refined crystal structure of the seryl-tRNA synthetase from Thermus thermophilus at 2.5 A resolution. J Mol Biol **234**(1): 222-33.

Furey, W. F. and Swaminathan, S. (1997). PHASES-95: A program package for processing and analyzing diffraction data from macromolecules. Methods Enzymol **277**: 590-620.

Gajiwala, K. S. and Burley, S. K. (2000a). Winged helix proteins. Curr Opin Struct Biol **10**(1): 110-6.

Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B. and Burley, S. K. (2000b). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature **403**(6772): 916-21.

Glasfeld, A., Koehler, A. N., Schumacher, M. A. and Brennan, R. G. (1999). The role of lysine 55 in determining the specificity of the purine repressor for its operators through minor groove interactions. J Mol Biol **291**(2): 347-61.

Glover, J. N. and Harrison, S. C. (1995). Crystal structure of the heterodimeric bZIP transcription factor c-Fos- c-Jun bound to DNA. Nature **373**(6511): 257-61.

Godsey, M. H., Baranova, N. N., Neyfakh, A. A. and Brennan, R. G. (2000). Crystallization and preliminary X-ray diffraction studies on the DNA- binding domain of the multidrug transporter activation protein (MtaN) from bacillus subtilis. Acta Crystallogr D Biol Crystallogr **56**(Pt 11): 1456-8.

Godsey, M. H., Baranova, N. N., Neyfakh, A. A. and Brennan, R. G. (2001). Crystal Structure of MtaN, a global multidrug-transporter gene activator. J Biol Chem **In Press**(published online ahead of print as 10.1074/jbc.M105819200).

Grinius, L. L. and Goldberg, E. B. (1994). Bacterial multidrug resistance is due to a single membrane protein which functions as a drug pump. J Biol Chem **269**(47): 29998-30004.

Grkovic, S., Brown, M. H., Roberts, N. J., Paulsen, I. T. and Skurray, R. A. (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., Brown, M. H. and Skurray, R. A. (2001a). Transcriptional regulation of multidrug efflux pumps in bacteria. Semin Cell Dev Biol **12**(3): 225-37.

Grkovic, S., Brown, M. H. and Skurray, R. A. (2001b). Transcriptional regulation of multidrug efflux pumps in bacteria. Semin Cell Dev Biol: In Press.

Guex, N. and Peitsch, M. C. (1997). Swiss-Model and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis **18**: 2714-23.

Hankinson, O. (1995). The aryl hydrocarbon receptor complex. Annu Rev Pharmacol Toxicol **35**: 307-40.

Harker, D. (1956). The determinatoin of phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. Acta Crystallogr **9**: 1-9.

Harrison, S. C. and Aggarwal, A. K. (1990). DNA recognition by proteins with the helix-turn-helix motif. Annu Rev Biochem **59**: 933-69.

Hayes, J. D. and Wolf, R. C. (1990). Molecular mechanisms of drug resistance. Biochem J **272**: 281-295.

Heldwein, E. E. and Brennan, R. G. (2001). Crystal structure of the transcription activator BmrR bound to DNA and a drug. Nature **409**(6818): 378-82.

Helmann, J. D. (1995). Compilation and analysis of Bacillus subtilis sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. Nucleic Acids Res **23**(13): 2351-60.

Helmann, J. D., Ballard, B. T. and Walsh, C. T. (1990). The MerR metalloregulatory protein binds mercuric ion as a tricoordinate, metal-bridged dimer. Science **247**(4945): 946-8.

Heltzel, A., Gambill, D., Jackson, W. J., Totis, P. A. and Summers, A. O. (1987). Overexpression and DNA-binding properties of the mer-encoded regulatory protein from plasmid NR1 (Tn21). J Bacteriol **169**(7): 3379-84.

Heltzel, A., Lee, I. W., Totis, P. A. and Summers, A. O. (1990). Activator-dependent preinduction binding of sigma-70 RNA polymerase at the metal-regulated mer promoter. Biochemistry **29**(41): 9572-84.

Hendrickson, W. A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science **254**(5028): 51-8.

Hendrickson, W. A., Horton, J. R. and LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. Embo J **9**(5): 1665-72.

Heymann, J. A., Sarker, R., Hirai, T., Shi, D., Milne, J. L., Maloney, P. C. and Subramaniam, S. (2001). Projection structure and molecular architecture of OxlT, a bacterial membrane transporter. Embo J **20**(16): 4408-4413.

Hidalgo, E. and Demple, B. (1997a). Spacing of promoter elements regulates the basal expression of the soxS gene and converts SoxR from a transcriptional activator into a repressor. Embo J **16**(5): 1056-65.

Hidalgo, E., Ding, H. and Demple, B. (1997b). Redox signal transduction via iron-sulfur clusters in the SoxR transcription activator. Trends Biochem Sci **22**(6): 207-10.

Higgins, C. F. and Gottesman, M. M. (1992). Is the multidrug transporter a flippase? Trends Biochem Sci **17**(1): 18-21.

Higgins, C. F., Hiles, I. D., Salmond, G. P., Gill, D. R., Downie, J. A., Evans, I. J., Holland, I. B., Gray, L., Buckel, S. D., Bell, A. W. and *et al.* (1986). A family of

related ATP-binding subunits coupled to many distinct biological processes in bacteria. Nature **323**(6087): 448-50.

Higgins, C. F., Hiles, I. D., Whalley, K. and Jamieson, D. J. (1985). Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems. Embo J **4**(4): 1033-9.

Holmes, D. J., Caso, J. L. and Thompson, C. J. (1993). Autogenous transcriptional activation of a thiostrepton-induced gene in Streptomyces lividans. Embo J **12**(8): 3183-91.

Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P. and Tainer, J. A. (2000). Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. Cell **101**(7): 789-800.

Hung, L. W., Wang, I. X., Nikaido, K., Liu, P. Q., Ames, G. F. and Kim, S. H. (1998). Crystal structure of the ATP-binding subunit of an ABC transporter. Nature **396**(6712): 703-7.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Bio **3**: 318-356.

Jair, K. W., Yu, X., Skarstad, K., Thony, B., Fujita, N., Ishihama, A. and Wolf, R. E., Jr. (1996). Transcriptional activation of promoters of the superoxide and multiple antibiotic resistance regulons by Rob, a binding protein of the Escherichia coli origin of chromosomal replication. J Bacteriol **178**(9): 2507-13.

Janin, J. (1995). Protein-protein recognition. Prog Biophys Mol Biol **64**(2-3): 145-66.

Janin, J. (1999). Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. Structure Fold Des **7**(12): R277-9.

Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. and Lecomte, C. (2000). Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. Proc Natl Acad Sci U S A **97**(7): 3171-6.

Jin, C. and Liao, X. (1999a). Backbone dynamics of a winged helix protein and its DNA complex at different temperatures: changes of internal motions in genesis upon binding to DNA. J Mol Biol **292**(3): 641-51.

Jin, C., Marsden, I., Chen, X. and Liao, X. (1999b). Dynamic DNA contacts observed in the NMR structure of winged helix protein-DNA complex. J Mol Biol **289**(4): 683-90.

Jones, S. and Thornton, J. M. (1996). Principles of protein-protein interactions. Proc Natl Acad Sci U S A **93**(1): 13-20.

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr A **47**(Pt 2): 110-9.

Juliano, R. L. and Ling, V. (1976). A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim Biophys Acta **455**(1): 152-62.

Kartner, N., Riordan, J. R. and Ling, V. (1983a). Cell surface P-glycoprotein associated with multidrug resistance in mammalian cell lines. Science **221**(4617): 1285-8.

Kartner, N., Shales, M., Riordan, J. R. and Ling, V. (1983b). Daunorubicin-resistant Chinese hamster ovary cells expressing multidrug resistance and a cell-surface P-glycoprotein. Cancer Res **43**(9): 4413-9.

Ke, H. (1997). Overview of isomorphous replacement phasing. New York, Academic Press.

Kessel, D. (1986). Circumvention of resistance to anthracyclines by calcium antagonists and other membrane-perturbing agents. Cancer Surv **5**(1): 109-27.

Kissinger, C. R., Gehlhaar, D. K. and Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. Acta Crystallogr D Biol Crystallogr **55**(Pt 2): 484-91.

Koronakis, V., Li, J., Koronakis, E. and Stauffer, K. (1997). Structure of TolC, the outer membrane component of the bacterial type I efflux system, derived from two-dimensional crystals. Mol Microbiol **23**(3): 617-26.

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. and Hughes, C. (2000). Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature **405**(6789): 914-9.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Danchin, A. and *et al.* (1997). The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature **390**(6657): 249-56.

Kwon, H. J., Bennik, M. H., Demple, B. and Ellenberger, T. (2000). Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA. Nat Struct Biol **7**(5): 424-30.

Laity, J. H., Lee, B. M. and Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol **11**(1): 39-46.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst **26**: 283-291.

Leigh, J. B. (1976). Synchrotron x-ray sources: a new tool in biological structural and kinetic analysis. Annu Rev Biophys Bioeng **5**: 239-70.

Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography **26**.

Ling, V., Kartner, N., Sudo, T., Siminovitch, L. and Riordan, J. R. (1983). Multidrug-resistance phenotype in Chinese hamster ovary cells. Cancer Treat Rep **67**(10): 869-74.

Linton, K. J. and Higgins, C. F. (1998). The Escherichia coli ATP-binding cassette (ABC) proteins. Mol Microbiol **28**(1): 5-13.

Livingstone, C. D. and Barton, G. J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput Appl Biosci **9**(6): 745-56.

Loh, J., Stacey, M. G., Sadowsky, M. J. and Stacey, G. (1999). The Bradyrhizobium japonicum nolA gene encodes three functionally distinct proteins. J Bacteriol **181**(5): 1544-54.

Lomovskaya, O., Lewis, K. and Matin, A. (1995). EmrR is a negative regulator of the Escherichia coli multidrug resistance pump EmrAB. J Bacteriol **177**(9): 2328-34.

Lowe, J., Cordell, S. C. and van den Ent, F. (2001). Crystal Structure of the SMC Head Domain: An ABC ATPase with 900 Residues Antiparallel Coiled-coil Inserted. J Mol Biol **306**(1): 25-35.

Lundblad, J. R., Laurance, M. and Goodman, R. H. (1996). Fluorescence polarization analysis of protein-DNA and protein-protein interactions. Mol Endocrinol **10**(6): 607-12.

Lunin, V. Y. (1988). Use of the information on electron density distribution in macromolecules. Acta Crystallogr A **44**(2): 144-50.

Lupas, A. (1997). Predicting coiled-coil regions in proteins. Curr Opin Struct Biol **7**(3): 388-93.

Lyon, B. R. and Skurray, R. (1987). Antimicrobial resistance of Staphylococcus aureus: genetic basis. Microbiol Rev **51**(1): 88-134.

Ma, D., Alberti, M., Lynch, C., Nikaido, H. and Hearst, J. E. (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.

Markham, P. N., Ahmed, M. and Neyfakh, A. A. (1996). The drug-binding activity of the multidrug-responding transcriptional regulator BmrR resides in its C-terminal domain. J Bacteriol **178**(5): 1473-5.

Martin, R. G., Jair, K. W., Wolf, R. E., Jr. and Rosner, J. L. (1996). Autoactivation of the marRAB multiple antibiotic resistance operon by the MarA transcriptional activator in Escherichia coli. J Bacteriol **178**(8): 2216-23.

Martin, R. G. and Rosner, J. L. (2001). The AraC transcriptional activators. Curr Opin Microbiol **4**(2): 132-7.

Matthews, B. W. (1968). Solvent content of protein crystals. J Mol Bio **33**: 491-497.

McClure, W. R. (1985). Mechanism and control of transcription initiation in prokaryotes. Annu Rev Biochem **54**: 171-204.

McKay, D. B. and Steitz, T. A. (1981). Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA. Nature **290**(5809): 744-9.

McPherson, A. (1999). Crystallization of biological macromolecules. Cold Spring Harbor, N.Y., Cold Spring Harbor Press.

McRee, D. E. (1999). Practical protein crystallography. San Diego, Academic Press.

Mine, T., Morita, Y., Kataoka, A., Mizushima, T. and Tsuchiya, T. (1999). Expression in Escherichia coli of a new multidrug efflux pump, MexXY, from Pseudomonas aeruginosa. Antimicrob Agents Chemother **43**(2): 415-7.

Mitchell, B. A., Brown, M. H. and Skurray, R. A. (1998). QacA multidrug efflux pump from Staphylococcus aureus: comparative analysis of resistance to diamidines,

biguanidines, and guanylhydrazones. Antimicrob Agents Chemother **42**(2): 475-7.

Monera, O. D., Kay, C. M. and Hodges, R. S. (1994). Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. Biochemistry **33**(13): 3862-71.

Mordoch, S. S., Granot, D., Lebendiker, M. and Schuldiner, S. (1999). Scanning cysteine accessibility of EmrE, an H+-coupled multidrug transporter from Escherichia coli, reveals a hydrophobic pathway for solutes. J Biol Chem **274**(27): 19480-6.

Muller, M. and Schimz, K. L. (1999). Oxazolidinones: a novel class of antibiotics. Cell Mol Life Sci **56**(3-4): 280-5.

Muth, T. R. and Schuldiner, S. (2000). A membrane-embedded glutamate is required for ligand binding to the multidrug transporter EmrE. Embo J **19**(2): 234-40.

Nikaido, H. (1996). Multidrug efflux pumps of gram-negative bacteria. J Bacteriol **178**(20): 5853-9.

Nikaido, H. (2001). Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria. Semin Cell Dev Biol **12**(3): 215-23.

Nikaido, H. and Zgurskaya, H. I. (2001). AcrAB and related multidrug efflux pumps of Escherichia coli. J Mol Microbiol Biotechnol **3**(2): 215-8.

Nishikawa, K. and Scheraga, H. A. (1976). Geometrical criteria for formation of coiled-coil structures of polypeptide chains. Macromolecules **9**(3): 395-407.

Noll, M., Petrukhin, K. and Lutsenko, S. (1998). Identification of a novel transcription regulator from Proteus mirabilis, PMTR, revealed a possible role of YJAI protein in balancing zinc in Escherichia coli. J Biol Chem **273**(33): 21393-401.

Nunberg, J. H., Kaufman, R. J., Schimke, R. T., Urlaub, G. and Chasin, L. A. (1978). Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate- resistant Chinese hamster ovary cell line. Proc Natl Acad Sci U S A **75**(11): 5553-6.

Outten, C. E., Outten, F. W. and O'Halloran, T. V. (1999). DNA distortion mechanism for transcriptional activation by ZntR, a Zn(II)-responsive MerR homologue in Escherichia coli. J Biol Chem **274**(53): 37517-24.

Outten, F. W., Outten, C. E., Hale, J. and O'Halloran, T. V. (2000). Transcriptional activation of an escherichia coli copper efflux regulon by the chromosomal MerR homologue, cueR. J Biol Chem **275**(40): 31024-9.

Pabo, C. O. and Lewis, M. (1982). The operator-binding domain of lambda repressor: structure and DNA recognition. Nature **298**(5873): 443-7.

Pannu, N. S. and Read, R. J. (1996). Improved structure refinement through maximum likelihood. Acta Crystallogr A **52**: 659-668.

Pao, S. S., Paulsen, I. T. and Saier, M. H., Jr. (1998). Major facilitator superfamily. Microbiol Mol Biol Rev **62**(1): 1-34.

Parkhill, J., Ansari, A. Z., Wright, J. G., Brown, N. L. and O'Halloran, T. V. (1993). Construction and characterization of a mercury-independent MerR activator (MerRAC): transcriptional activation in the absence of Hg(II) is accompanied by DNA distortion. Embo J **12**(2): 413-21.

Parkhill, J. and Brown, N. L. (1990). Site-specific insertion and deletion mutants in the mer promoter- operator region of Tn501; the nineteen base-pair spacer is essential for normal induction of the promoter by MerR. Nucleic Acids Res **18**(17): 5157-62.

Parkhill, J., Lawley, B., Hobman, J. L. and Brown, N. L. (1998). Selection and characterization of mercury-independent activation mutants of the Tn501 transcriptional regulator, MerR. Microbiology **144**(Pt 10): 2855-64.

Paulsen, I. T., Brown, M. H. and Skurray, R. A. (1996). Proton-dependent multidrug efflux systems. Microbiol Rev **60**(4): 575-608.

Phillips, J. C. and Hodgson, K. O. (1980). the use of anomalous scattering effects to phase diffraction patterns from macromolecules. Acta Crystallogr A **36**: 856-864.

Phillips, S. E. (1994). Built by association: structure and function of helix-loop-helix DNA- binding proteins. Structure **2**(1): 1-4.

Putman, M., van Veen, H. W. and Konings, W. N. (2000). Molecular properties of bacterial multidrug transporters. Microbiol Mol Biol Rev **64**(4): 672-93.

Rabussay, D. and Zellig, W. (1969). FEBS letters **5**: 104-106.

Ralston, D. M. and O'Halloran, T. V. (1990). Ultrasensitivity and heavy-metal selectivity of the allosterically modulated MerR transcription complex. Proc Natl Acad Sci U S A **87**(10): 3846-50.

Ramakrishnan, V. and Biou, V. (1997). Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement. Methods Enzymol **276**: 538-57.

Raviv, Y., Pollard, H. B., Bruggemann, E. P., Pastan, I. and Gottesman, M. M. (1990). Photosensitized labeling of a functional multidrug transporter in living drug-resistant tumor cells. J Biol Chem **265**(7): 3975-80.

Reynolds, P. E. (1984). Resistance of the antibiotic target site. Br Med Bull **40**(1): 3-10.

Rhee, S., Martin, R. G., Rosner, J. L. and Davies, D. R. (1998). A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. Proc Natl Acad Sci U S A **95**(18): 10413-8.

Rhodes, G. (1993). Crystallography made crystal clear. San Diego, Academic Press.

Roninson, I. B., Chin, J. E., Choi, K. G., Gros, P., Housman, D. E., Fojo, A., Shen, D. W., Gottesman, M. M. and Pastan, I. (1986). Isolation of human mdr DNA sequences amplified in multidrug-resistant KB carcinoma cells. Proc Natl Acad Sci U S A **83**(12): 4538-42.

Rosenberg, M. F., Callaghan, R., Ford, R. C. and Higgins, C. F. (1997). Structure of the multidrug resistance P-glycoprotein to 2.5 nm resolution determined by electron microscopy and image analysis. J Biol Chem **272**(16): 10685-94.

Rosenberg, M. F., Mao, Q., Holzenburg, A., Ford, R. C., Deeley, R. G. and Cole, S. P. (2001). The structure of the multidrug resistance protein 1 (MRP1/ABCC1). crystallization and single-particle analysis. J Biol Chem **276**(19): 16076-82.

Rossmann, M. G. (1990). The molecular replacement method. Acta Crystallogr A **46**(Pt 2): 73-82.

Sadowsky, M. J., Cregan, P. B., Gottfert, M., Sharma, A., Gerhold, D., Rodriguez-Quinones, F., Keyser, H. H., Hennecke, H. and Stacey, G. (1991). The

Bradyrhizobium japonicum nolA gene and its involvement in the genotype-specific nodulation of soybeans. Proc Natl Acad Sci U S A **88**(2): 637-41.

Saier, M. H., Jr., Beatty, J. T., Goffeau, A., Harley, K. T., Heijne, W. H., Huang, S. C., Jack, D. L., Jahn, P. S., Lew, K., Liu, J., Pao, S. S., Paulsen, I. T., Tseng, T. T. and Virk, P. S. (1999). The major facilitator superfamily. J Mol Microbiol Biotechnol **1**(2): 257-79.

Saier, M. H., Jr. and Paulsen, I. T. (2001). Phylogeny of multidrug transporters. Semin Cell Dev Biol **12**(3): 205-13.

Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. and Pabo, C. O. (1982). Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature **298**(5873): 447-51.

Schneider, E. and Hunke, S. (1998). ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. FEMS Microbiol Rev **22**(1): 1-20.

Schuldiner, S., Granot, D., Steiner, S., Ninio, S., Rotem, D., Soskin, M. and Yerushalmi, H. (2001). Precious things come in little packages. J Mol Microbiol Biotechnol **3**(2): 155-62.

Schuldiner, S., Lebendiker, M. and Yerushalmi, H. (1997). EmrE, the smallest ion-coupled transporter, provides a unique paradigm for structure-function studies. J Exp Biol **200**(Pt 2): 335-41.

Schumacher, M. A., Choi, K. Y., Lu, F., Zalkin, H. and Brennan, R. G. (1995). Mechanism of corepressor-mediated specific DNA binding by the purine repressor. Cell **83**(1): 147-55.

Schumacher, M. A., Choi, K. Y., Zalkin, H. and Brennan, R. G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. Science **266**(5186): 763-70.

Schwaiger, M., Lebendiker, M., Yerushalmi, H., Coles, M., Groger, A., Schwarz, C., Schuldiner, S. and Kessler, H. (1998). NMR investigation of the multidrug transporter EmrE, an integral membrane protein. Eur J Biochem **254**(3): 610-9.

Seeman, N. C., Rosenberg, J. M. and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. Proc Natl Acad Sci U S A **73**(3): 804-8.

Shen, D. W., Fojo, A., Chin, J. E., Roninson, I. B., Richert, N., Pastan, I. and Gottesman, M. M. (1986). Human multidrug-resistant cell lines: increased mdr1 expression can precede gene amplification. Science **232**(4750): 643-5.

Shewchuk, L. M., Helmann, J. D., Ross, W., Park, S. J., Summers, A. O. and Walsh, C. T. (1989a). Transcriptional switching by the MerR protein: activation and repression mutants implicate distinct DNA and mercury(II) binding domains. Biochemistry **28**(5): 2340-4.

Shewchuk, L. M., Verdine, G. L., Nash, H. and Walsh, C. T. (1989b). Mutagenesis of the cysteines in the metalloregulatory protein MerR indicates that a metal-bridged dimer activates transcription. Biochemistry **28**(15): 6140-5.

Smith, J. T. and Amyes, S. G. (1984). Bacterial resistance to antifolate chemotherapeutic agents mediated by plasmids. Br Med Bull **40**(1): 42-6.

Stark, G. R. (1986a). Cancer chemotherapy. Progress in understanding multidrug resistance. Nature **324**(6096): 407-8.

Stark, G. R. (1986b). DNA amplification in drug resistant cells and in tumours. Cancer Surv **5**(1): 1-23.

Steitz, T. A. (1990). Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. Q Rev Biophys **23**(3): 205-80.

Stout, G. H. and Jensen, L. H. (1989). X-ray structure determination: a practical guide. New York, John Wiley & Sons.

Summers, A. O. (1992). Untwist and shout: a heavy metal-responsive transcriptional regulator. J Bacteriol **174**(10): 3097-101.

Tate, C. G., Kunji, E. R. S., Lebendiker, M. and Schuldiner, S. (2001). The projection structure of EmrE, a proton-linked multidrug transporter from *Eschericia coli*, at 7 Å resolution. Embo J **20**: 77-81.

Tennent, J. M., Lyon, B. R., Midgley, M., Jones, I. G., Purewal, A. S. and Skurray, R. A. (1989). Physical and biochemical characterization of the qacA gene encoding antiseptic and disinfectant resistance in Staphylococcus aureus. J Gen Microbiol **135**(Pt 1): 1-10.

Tenover, F. C. (1999). Implications of vancomycin-resistant Staphylococcus aureus. J Hosp Infect **43 Suppl**: S3-7.

Terwilliger, T. C. and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr **55**(Pt 4): 849-61.

Thoden, J. B., Ruzicka, F. J., Frey, P. A., Rayment, I. and Holden, H. M. (1997). Structural analysis of the H166G site-directed mutant of galactose-1- phosphate uridylyltransferase complexed with either UDP-glucose or UDP- galactose:

detailed description of the nucleotide sugar binding site. Biochemistry **36**(6): 1212-22.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22**(22): 4673-80.

Travis, J. (1994). Reviving the antibiotic miracle? Science **264**(5157): 360-2.

Ueda, K., Taguchi, Y. and Morishima, M. (1997). How does P-glycoprotein recognize its substrates? Semin Cancer Biol **8**(3): 151-9.

Utschig, L. M., Bryson, J. W. and O'Halloran, T. V. (1995). Mercury-199 NMR of the metal receptor site in MerR and its protein-DNA complex. Science **268**(5209): 380-5.

van den Ent, F., Lockhart, A., Kendrick-Jones, J. and Lowe, J. (1999). Crystal structure of the N-terminal domain of MukB: a protein involved in chromosome partitioning. Structure Fold Des **7**(10): 1181-7.

Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J Mol Biol **229**(1): 105-24.

van Veen, H. W. (2001). Towards the molecular mechanism of prokaryotic and eukaryotic multidrug transporters. Semin Cell Dev Biol **12**(3): 239-45.

van Veen, H. W., Margolles, A., Muller, M., Higgins, C. F. and Konings, W. N. (2000). The homodimeric ATP-binding cassette transporter LmrA mediates multidrug

transport by an alternating two-site (two-cylinder engine) mechanism. Embo J **19**(11): 2503-14.

van Veen, H. W., Putman, M., Margolles, A., Sakamoto, K. and Konings, W. N. (1999). Structure-function analysis of multidrug transporters in Lactococcus lactis. Biochim Biophys Acta **1461**(2): 201-6.

Vasudevan, G., Ullman, B. and Landfear, S. M. (2001). Point mutations in a nucleoside transporter gene from Leishmania donovani confer drug resistance and alter substrate selectivity. Proc Natl Acad Sci U S A **98**(11): 6092-7.

Vazquez-Laslop, N., Markham, P. N. and Neyfakh, A. A. (1999). Mechanism of ligand recognition by BmrR, the multidrug-responding transcriptional regulator: mutational analysis of the ligand-binding site. Biochemistry **38**(51): 16925-31.

Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. and Sussman, J. L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. Proc Natl Acad Sci U S A **97**(2): 623-8.

Wolberger, C. and Campbell, R. (2000). New perch for the winged helix. Nat Struct Biol **7**(4): 261-2.

Wolf, E., Kim, P. S. and Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci **6**(6): 1179-89.

Woolridge, D. P., Martinez, J. D., Stringer, D. E. and Gerner, E. W. (1999). Characterization of a novel spermidine/spermine acetyltransferase, BltD, from Bacillus subtilis. Biochem J **340**(Pt 3): 753-8.

Woolridge, D. P., Vazquez-Laslop, N., Markham, P. N., Chevalier, M. S., Gerner, E. W. and Neyfakh, A. A. (1997). Efflux of the natural polyamine spermidine facilitated by the Bacillus subtilis multidrug transporter Blt. J Biol Chem **272**(14): 8864-6.

Yerushalmi, H., Lebendiker, M. and Schuldiner, S. (1995). EmrE, an Escherichia coli 12-kDa multidrug transporter, exchanges toxic cations and H+ and is soluble in organic solvents. J Biol Chem **270**(12): 6856-63.

Yerushalmi, H. and Schuldiner, S. (2000). An essential glutamyl residue in EmrE, a multidrug antiporter from Escherichia coli. J Biol Chem **275**(8): 5264-9.

Zeng, Q., Stalhandske, C., Anderson, M. C., Scott, R. A. and Summers, A. O. (1998). The core metal-recognition domain of MerR. Biochemistry **37**(45): 15885-95.

Zheleznova, E. E., Markham, P., Edgar, R., Bibi, E., Neyfakh, A. A. and Brennan, R. G. (2000). A structure-based mechanism for drug binding by multidrug transporters. Trends Biochem Sci **25**(2): 39-43.

Zheleznova, E. E., Markham, P. N., Neyfakh, A. A. and Brennan, R. G. (1997). Preliminary structural studies on the multi-ligand-binding domain of the transcription activator, BmrR, from Bacillus subtilis. Protein Sci **6**(11): 2465-8.

Zheleznova, E. E., Markham, P. N., Neyfakh, A. A. and Brennan, R. G. (1999). Structural basis of multidrug recognition by BmrR, a transcription activator of a multidrug transporter. Cell **96**(3): 353-62.

Zheleznova-Heldwein, E. E. and Brennan, R. G. (2001). Crystal structure of the transcription activator BmrR bound to DNA and a drug. Nature **409**(6818): 378-82.

Zhou, N. E., Kay, C. M. and Hodges, R. S. (1994). The role of interhelical ionic interactions in controlling protein folding and stability. De novo designed synthetic two-stranded alpha- helical coiled-coils. J Mol Biol **237**(4): 500-12.

Zhou, Z., White, K. A., Polissi, A., Georgopoulos, C. and Raetz, C. R. (1998). Function of Escherichia coli MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. J Biol Chem **273**(20): 12466-75.