CRYSTALLOGRAPHIC STUDIES ON THE MULTIDRUG-BINDING TRANSCRIPTION ACTIVATOR BMRR, FROM *B.SUBTILIS*, AND *E.COLI* PURINE REPRESSOR

by

Ekaterina E. Zheleznova

A DISSERTATION

Presented to the Department of Biochemistry and Molecular Biology
and the Oregon Health Sciences University
School of Medicine
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

March, 1999

School of Medicine

Oregon Health Sciences University

CERTIFICATE OF APPROVAL

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

Ekaterina E. Zheleznova

has been approved

[redacted]

Professor in charge of thesis

[redacted]

Member

[redacted]

Member

[redacted]

Member

[redacted]

Member

[redacted]

Associate Dean for Graduate Studies

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor, Dr. Richard Brennan, for luring me into the fascinating world of crystallography, for providing me with excellent training as well as financial support, and for sparing neither encouragement nor criticism during my graduate work.

I would certainly like to acknowledge our collaborators on the BRC project, Dr. Alex Neyfakh and Dr. Penelope Markham from the University of Illinois at Chicago, for providing me with purified BRC protein and making mutant proteins.

I would also like to thank all of my committee members for graciously agreeing to read my thesis and evaluate my graduate work: Dr. Scott Landfear, Dr. Dave Farrens, and Dr. Hans Peter Bächinger; Dr. Bächinger for his biophysical expertise, which brings order into the chaotic biological world.

I am grateful to all my friends and colleagues for making my American experience a blast and for being within reach at times of desperation. I particularly want to acknowledge a superb scientist, Dr. Armando Jardim, for his help throughout the course of my graduate work. Among the members of the Brennan lab, I wish to thank my friend and isle-mate Joy Huffman for being herself and Dr. Maria Schumacher for setting an example that all graduate students should try to follow.

My special thanks are due to Dr. Jill Kelly, Kurt Heldwein, and Dr. Armando Jardim for proofreading parts of this dissertation and for helpful discussions.

Finally, I am indebted to my family for their endless love and support: especially, to my sister, Yulya; to my parents, Evgeny and Natalie, for their guidance and constant encouragement to pursue my dreams; and to my grandparents, Michail and Zoe, for setting impeccable standards I am trying to live up to. Last, but certainly not least, I want to thank Kurt Heldwein for being a terrific partner and for putting up with my "writing tantrums" in the last several months.

ABSTRACT

Multidrug efflux transporters are ubiquitous membrane-bound proteins that by yet unknown mechanisms can bind and remove from cells a variety of structurally dissimilar drugs and toxins. Among substrates of multidrug transporters are antitumor and antimicrobial agents, and therefore, understanding the mechanism of multidrug recognition has tremendous clinical importance.

A comparable ability to bind diverse hydrophobic cationic drugs is characteristic of the *Bacillus subtilis* transcription regulator BmrR, which upon binding a drug, activates expression of a multidrug transporter. To unveil the mechanism of multidrug recognition by BmrR, we determined the three-dimensional crystal structures of the multidrug-binding domain of BmrR (BRC) in its apo form as well as in complex with a drug, tetraphenylphosphonium (TPP) to 2.7 Å and 2.8 Å, respectively.

The crystallographic analysis of apo BRC and BRC-TPP complex revealed the following mechanism of ligand binding. TPP penetrates into the hydrophobic core of the protein, where it forms a number of van der Waals and stacking interactions with the surrounding hydrophobic residues, but most importantly, an electrostatic interaction with a buried glutamate residue, Glu134, which is the key to cation selectivity of BRC. In the apo structure, the drug-binding site is completely inaccessible, but in the presence of a drug, a helical "lid" of the binding site becomes disordered thus allowing access to the binding site. The two structures provide the first atomic view of a multidrug-binding site and outline the principles of multidrug recognition by BmrR. Moreover, they suggest a model by which multidrug transporters may recognize their cationic substrates.

The *E. coli* purine repressor PurR binds its cognate palindromic operators with high selectivity through a series of sequence-specific contacts. However, PurR makes no sequence-specific contacts to the conserved central CpG base-pair step, which is found in all binding sites. To understand the requirements imposed on this central sequence by PurR, we compared the chemical structures of CpG and TpA sequences and dissected the differences between the former and the latter down to specific exocyclic positions. Specifically, we designed a series of systematically altered *purF* operators, where CpI, TpA, UpD, TpD, UpAP, TpAP, 5meCpG, or UpiG steps replaced the central CpG sequence. Through a combination of structural and binding-affinity studies, we identified the key determinants of PurR specificity for the central sequence: a 5-methyl group on the pyrimidine and a 2-amino group on the purine. In the major groove, the 5-methyl group interferes with high-affinity binding because this group comes into an unfavourable steric apposition with the group at the exocyclic 6 position of a neighbouring purine of the PurR-bound kinked *purF* operator. This steric strain increases the energetic cost of attaining the proper kinked DNA conformation in PurR·DNA complex. In the minor groove, the 2-amino group on the purine is indispensable for binding because it likely assists the folding of the hinge helix by expanding the minor groove of the CpG sequence. Taken together, these results indicate dramatic yet indirect effect of the DNA sequence on PurR recognition properties.

CHAPTER 1

INTRODUCTION

Correct recognition is crucial to almost every cellular event. Recognition rules apply whether the participating molecules are large, such as proteins and nucleic acids, or small, such as ions. The specificity can be high or low depending on the function that recognition performs and how important that function is. Recognition is important in the cellular processes because it allows molecules to communicate and to transduce signals.

This dissertation presents two separate studies on macromolecular recognition: the first shows how the purine repressor, PurR, can specifically read the central sequence of its operator without any specific contacts while the second reveals how the bacterial multidrug resistance regulator, BmrR, recognizes multiple dissimilar drugs.

Most proteins are highly specific for their ligands or substrates because they have to recognize their proper substrates among many incorrect ones. The recognition closely parallels the process of reading. Proteins read their substrates, breaking them down into recognizable patterns and words, then reconstructing them back into whole sentences, just as readers do. As an example, DNA-binding proteins find and specifically bind their cognate sites among a myriad of others. They read different parts of the nucleotide sequence trying to figure out certain "correct" patterns. One may think of them reading each base pair as a word, independently from its neighbours, as well as reading the global conformation of the DNA as if trying to figure out whether the whole sentence makes sense. Some words–DNA bases–are more important then others for a given DNA-binding protein; in binding sites, they occur at so called conservative positions. Proteins, like readers, can tolerate a certain number of misspelled cues as long as they are able to read the whole message. One subject of this dissertation, an *E.coli* transcription regulator, PurR, recognizes its cognate sites with high specificity and, in particular, prefers the CpG

sequence in the middle of its binding site. Yet, the basis for such discrimination is puzzling because PurR does not make any specific protein-DNA base contacts with this CpG sequence. In the following study, we address the reason for such stringent discrimination.

Multidrug-binding proteins, such as multidrug transporters, expel various toxic compounds out of the cell, and they clearly violate the rules of specific recognition by binding a variety of molecules that have no obvious structural similarities. But multidrug transporters do not bind just any molecule; therefore, they must recognize some common pattern–perhaps a property of a drug–which is still unclear to us. It is almost as if multidrug-binding proteins are "reading" their substrates and are looking for a specific pattern they are programmed to recognize. The problem is that we do not understand what pattern they are looking for. The second subject of this dissertation, the *B.subtilis* multidrug-binding protein BmrR, recognizes and binds a variety of structurally and chemically unrelated molecules, multidrug transporters alike. Our current hypothesis is that BmrR is likely to filter its proper ligands by their physical properties rather than specific moieties. However, the intriguing polyspecific mechanism that BmrR employs is still unknown and is the subject of the second study presented below.

Understanding the specificity of biological macromolecules requires a detailed atomic-level knowledge of their structures. Crystallography is a powerful method that allows us to discern macromolecular structures at the level of atoms and to study the macromolecular recognition. X-ray crystallography is the primary method used in this work.

Ultimately, we would like to understand the rules that govern specificity of macromolecular recognition because it will allow us to understand the cellular processes in all their complexity and, perhaps, to influence them. This dissertation focuses on two different proteins that represent two distinct examples of macromolecular recognition.

A. Multidrug-binding proteins.

1.1. Multidrug transporters.

Multidrug transporters are ubiquitous membrane-bound proteins that actively remove multiple structurally-dissimilar compounds out of cells, thereby protecting those cells from the toxic effects of these chemicals. Since many anticancer, antibacterial, and antifungal agents used as front-line chemotherapeutics are substrates of multidrug transporters, these proteins seriously impede the successful treatment of cancer and infectious diseases. But clinical importance is not the only reason behind the interest in multidrug transporters. From a basic science viewpoint, these proteins are remarkable because of their extraordinary abilities to bind and transport a wide variety of structurally unrelated compounds.

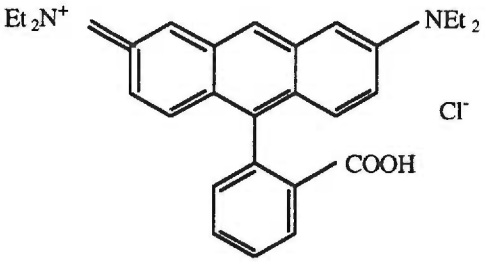
To date, over 100 multidrug transporters have been discovered both in prokaryotes and eukaryotes [Gottesman & Pastan, 1993; Lewis, 1994; Nikaido, 1994; Paulsen et al., 1996a]. However, the first multidrug transporter, mammalian phospho-glycoprotein (or P-glycoprotein), was identified over 20 years ago [Juliano and Ling, 1976; Riordan and Ling, 1979] as a protein overexpressed in multidrug-resistant tumor cells; subsequently, the *mdr* gene encoding P-glycoprotein was cloned. Moreover, through transfection of the *mdr* gene into multidrug-sensitive tumor cells and normal human cells, the P-glycoprotein was shown to be necessary and sufficient to impart a multidrug-resistant phenotype on these cells. Later, biochemical studies using purified P-glycoprotein showed that it indeed served as a transporter and that it had an unprecedented substrate specificity for a wide variety of structurally and chemically unrelated compounds. Thus, P-glycoprotein was found to be solely responsible for rendering tumor cells resistant to a broad spectrum of antitumor drugs and other toxins.

For over a decade, P-glycoprotein, remained the only known multidrug transporter [Gottesman & Pastan, 1993], but recently, several groups have identified multidrug

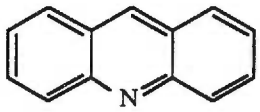
transporters in both Gram-negative and Gram-positive bacteria [Paulsen et al., 1996a; Lewis, 1994; Nikaido, 1994], yeast [Goffeau, 1997], and protozoa [Ullman, 1995]. Only Mycoplasma appears to lack these proteins [Fraser et al., 1995]. Sequence analysis has revealed that multidrug transporters belong to four different families of proteins: ATP-binding cassette (ABC) [Hyde at al, 1990], major facilitator (MRF) [Marger and Saier. 1993], small multidrug resistance (SMR) [Paulsen et al., 1996b], and resistance-nodulation (RND) [Saier et al., 1994]. The ATP-binding cassette family, represented by P-glycoprotein, uses ATP as an energy source to pump its substrates out of the cell while the other three families include proton-dependent transporters and derive the energy for drug efflux from the transmembrane proton gradient rather than from ATP hydrolysis.

All multidrug transporters efflux a wide variety of structurally unrelated organic compounds, including dyes, DNA-intercalators, antibiotics, anticancer drugs, etc. For example, multidrug transporter Bmr (**B**acterial **m**ultidrug **r**esistance) from bacteria *Bacillus subtilis* pumps a broad spectrum of compounds out of the cell, including ethidium bromide, tetraphenylphosphonium, puromycin, chloramphenicol, rhodamine and acridine dyes (Figure 1.1) as well as fluoroquinolone antibiotics [Neyfakh, et al., 1991]. Interestingly, many of these drugs are also substrates of P-glycoprotein, and drug efflux can be blocked by reserpine, a competitive inhibitor of P-glycoprotein.

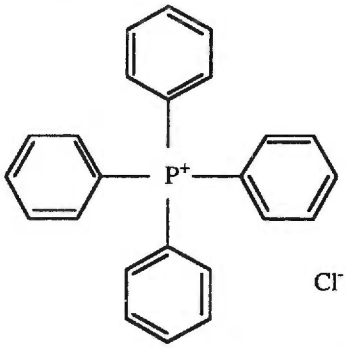
Suprisingly, even though multidrug transporters from different families are functionally similar, they do not share sequence homology or transmembrane topology. Often, they are more closely related to highly substrate-specific transporters. For example, the closest homologues of the *B. subtilis* transporter, Bmr, are tetracycline-specific transporters of Gram-negative bacteria [Neyfakh et al., 1991]. Analogously, P-glycoprotein is the only promiscuous protein in the ABC family of transporters: other members of this family are highly specific for their substrates, each preferring to bind and transport only one compound [Gottesman and Pastan, 1993]. Some theories suggest that in the course of evolution, multidrug transporters have likely arisen from specific drug-



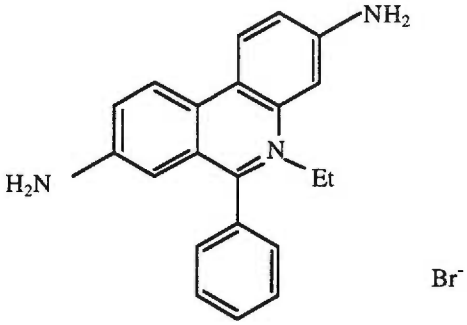
Rhodamine



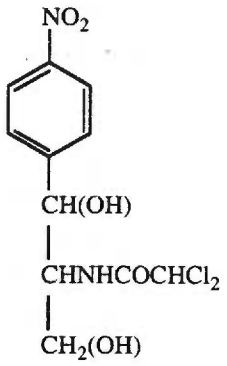
Acridine



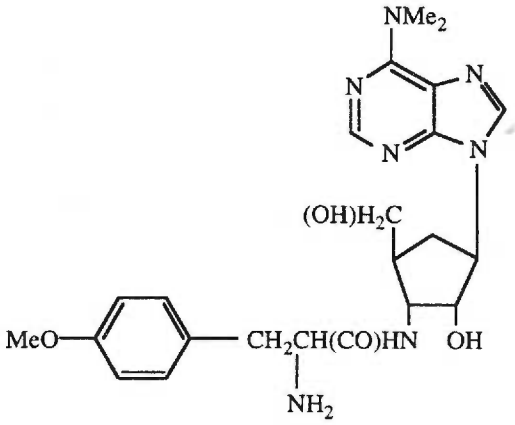
Tetraphenylphosphonium



Ethidium bromide



Chloramphenicol



Puromycin

Figure 1.1. Selected substrates of multidrug transporter Bmr from *Bacillus subtilis* [Neyfakh, et al., 1991].

extrusion pumps and not from a common multidrug-binding ancestor [Lewis, 1994; Neyfakh, 1997]. As an exception, some multidrug transporters within the RND family are closely related, and all 6 known members of the SMR family are highly homologous [Paulsen et al., 1996a].

Thus, multidrug transporters belong to different structural families of transporters, but their substrate spectra overlap. The challenge now is to identify the mechanism that these multidrug transporters utilize to recognize and transport structurally diverse compounds.

1.2. Models of multidrug binding and transport.

Since the discovery of the first multidrug transporter P-glycoprotein, scientists have been facing the challenge of explaining how these proteins are able to bind a plethora of chemically and structurally unrelated molecules. Indeed, the drug substrates of multidrug transporters share no structural similarity, except that the majority of them are lipophilic molecules carrying a positive electrostatic charge. Among the substrates of multidrug transporters are such dissimilar molecules as anticancer drugs (anthracyclines, Vinca alkaloids, taxanes), antibiotics (chloramphenicol and puromycin), quinolones, dyes (rhodamine and acridine), DNA intercalators (ethidium bromide), and steroids [Gros et al., 1992; Ito et al., 1997; Sharom, 1997]. The broad substrate specificity displayed by these transporters contrasts dramatically with the narrow chemical specificity of the vast majority of ligand-binding proteins. Intriguingly, these substrates can compete for binding, which suggests that each multidrug transporter has a single multidrug-binding site.

Logic tells us that the nature of the binding site of a multidrug transporter may be quite different from the enzyme-like sites that generally exhibit a high substrate-specificity. It is widely accepted that the three-dimensional shape of a protein determines its ability to specifically bind ligands; interactions between proteins and ligands demonstrate both steric and physical complementarity [Creighton, 1993]. Currently, two major binding theories

are employed to explain the substrate specificity of proteins. The "key-lock" model proposes that a ligand (key) fits precisely into the binding site of a protein (lock), and the structure of the protein doesn't change upon ligand binding. In the alternative "induced fit" model [Koshland, 1962], the presence of a substrate (or ligand) induces conformational changes at the binding site so as to fit the structure of the substrate and optimize its binding.

If we attempt to apply the first theory to explain multidrug recognition, then we have to imagine a rigid lock that could fit a variety of keys. With the second theory, which is more plausible, the multidrug transporter would possess an extremely flexible binding site, one that could be adjusted to accommodate a broad spectrum of structurally unrelated molecules.

To circumvent the paradox of multidrug recognition, several hypotheses have proposed that mdr transporters promote the efflux of various compounds without direct binding. One such hypothesis suggested that P-glycoprotein alters the drug distribution indirectly by affecting the membrane potential [Roepe et al., 1993]. Another popular hypothesis describes P-glycoprotein as a "hydrophobic vacuum cleaner", which indiscriminately expels drugs out of the membrane without making specific molecular contacts [Gottesman and Pastan, 1993]. This latter model is based on observations that many drugs are removed directly from the lipid bilayer and not from the cytoplasm as was originally thought [Homolya et al., 1993; Bolhuis et al., 1996; Shapiro et al., 1997]. The "vacuum cleaner" model can explain the unusually broad specificity of P-glycoprotein because the specificity is primarily determined by the drug's ability to intercalate into the lipid bilayer (its hydrophobicity) and not by interaction with a binding site.

However, abundant evidence imply that multidrug transporters directly bind their substrates and inhibitors. First, a large number of mutants of P-glycoprotein [Germann, 1996; Hanna et al., 1996] and several other multidrug transporters such as QacA [Paulsen et al., 1996c], Smr [Paulsen et al., 1996b], MdfA [Edgar and Bibi, 1999], and Bmr [Ahmed et al., 1993; Klyachko et al., 1997] display altered transport characteristics and

affinities for specific substrates and inhibitors. These point mutations indicate that the spectrum of substrates depends not only on substrate characteristics (hydrophobicity and charge) but also on the specific sequence and three-dimensional structure of the protein. Interestingly, mutational analysis of several multidrug transporters has shown that some acidic residues are critical to binding and essential to multidrug transport. For example, a glutamate in the first transmembrane domain is conserved among all multidrug transporters of the small multidrug resistance (SMR) family. This glutamate E13 is essential for transport activity of Smr protein, because substitution to any other residues (even aspartate) abolish transport [Paulsen et al., 1996b]. Similarly, in the *S. aureus* multidrug transporter QacA, a transmembrane acidic residue at position 322 or 323 is required for substrate binding [Paulsen et al., 1996c]. In an *E. coli* multidrug transporter, MdfA, the only intramembrane negatively-charged residue E26 is also essential for substrate binding, although not for transport.

Second, photoreactive analogues of many substrates and inhibitors of P-glycoprotein including vinblastine [Cornwell et al., 1986], azidopin [Greenberger et al., 1991; Bruggemann et al., 1992], forskolin [Morris et al., 1991], etc. have been shown to bind directly to P-glycoprotein upon irradiation. Such photoaffinity labelling is specific because it can be blocked by an excess of competing substrates of P-glycoprotein [Morris et al., 1991] or inhibitors such as verapamil [Cornwell et al., 1986]. Moreover, photoaffinity studies identified major sites of labelling in transmembrane segments 5, 6, 11, and 12 [Greenberger et al., 1991; Germann, 1996; Loo and Clarke, 1996].

Third, biophysical techniques have provided evidence for substrate-induced conformational changes in P-glycoprotein. Specifically, fluorescence techniques using highly purified P-glycoprotein showed that it undergoes significant conformational changes upon binding drugs [Liu and Sharom, 1996] and that the drug binding site is conformationally coupled to the ATP-binding domains [Liu and Sharom, 1996]. Also, an infrared spectroscopic study has reported a change in the tertiary structure of P-

glycoprotein upon binding verapamil and ATP [Sonveaux et al., 1996]. Finally, multidrug transporters do not efflux just any lipophilic molecule and rather are polyspecific.

Thus, the specific structure of the transporter along with the properties of substrates themselves dictates the spectrum of substrates and inhibitors. Consequently, multidrug transporters must function in a somewhat traditional way: they bind their substrates within a substrate-recognition site and only then expel them out of cells. Only, unlike traditional transporters, multidrug transporters possess remarkably promiscuous recognition sites, sites that can interact, in an unexplained way, with a large spectrum of dissimilar molecules.

1.3. The flippase model of a multidrug transporter.

The observation that multidrug transporters expel drugs out of cells into the extracellular space produced one of the first models depicting these transporters as classical pumps [Gottesman and Pastan, 1988]. These membrane proteins are thought to alternate between the inward-facing and outward-facing conformation. Pumps recruit their substrates in the cytoplasm, bind them in a specific site, and then undergo a conformational change to release the substrate into the extracellular space. Transmembrane domains of such a pump form a pore-like structure through which the drug passes without contacting the membrane lipids.

However, evidence accumulated to date suggests that multidrug transporters may act more like flippases. By definition, a flippase moves its substrates from the inner to the outer leaflet of a cellular membrane, instead of pumping molecules from the cytosol into the extracellular space. As a net result, a substrate is removed from the cytosol. According to this model, a multidrug transporter directly binds its nonpolar substrates within the cytoplasmic leaflet of the membrane and then 'flips' drugs to the exoplasmic leaflet of the bilayer or directly into the extracellular medium [Higgins and Gottesman, 1992]. Indeed, a typical multidrug-transporter substrate is a relatively hydrophobic, positively-charged

molecule containing aromatic rings and can intercalate between phospholipid groups in the lipid bilayer [Gros et al., 1992; Sharom, 1997].

Substantial experimental evidence supports the view that substrates interact with multidrug transporters within the membrane. On of the first indications came from experiments with 5'[^{125}I]iodonaphtalene-1-azide (INA), a lipid-soluble agent which labels membrane proteins when photoactivated. In multidrug-resistant cells, INA becomes photoactivated by energy transfer from P-glycoprotein substrates, doxorubicin and rhodamine, and labels the P-glycoprotein *in vivo* [Raviv et al., 1990].

Furthermore, studies on the P-glycoprotein using hydrophobic acetomethyl derivatives of fluorescent dyes showed that many substrates of P-glycoprotein are expelled directly from the cellular membrane [Homolya et al., 1993]. If a nonfluorescent acetomethyl derivative gains access to the cytosol, it is rapidly hydrolyzed by cytosolic esterases to the highly fluorescent free-acid form. The free-acid form is not transported by the P-glycoprotein and thus, remains trapped in the cytosol. However, in multidrug-resistant cells, the free-acid form does not accumulate; presumably, the derivative compound is expelled by the transporter before the former has a chance to reach the cytosol. An analogous study done on multidrug transporter LmrA from bacteria *Lactococcus lactis* has yielded very similar results [Bolhuis et al., 1996].

Yet another example of the intramembrane substrate recognition is the distribution of P-glycoprotein substrate Hoechst 33342 in multidrug-resistant vs. multidrug-sensitive cells. Hoechst 33342 is only fluorescent in a hydrophobic environment. In multidrug-sensitive cells, this dye accumulates in the hydrophobic environment of the lipid bilayer whereas in multidrug-resistant cells, Hoechst 33342 is expelled out of the cell membranes by P-glycoprotein and is mostly found in aqueous phase [Shapiro et al., 1997]. Finally, van Helvoort et al. discovered that P-glycoprotein is indeed able to translocate short-chained derivatives of various membrane lipids from the inner to the outer leaflet of epithelial cells, a hallmark of a "flippase" function [1996].

All these data may lead one to a conclusion that multidrug transporters, in particular, P-glycoprotein, transport their substrates out of the lipid bilayer. However, to prove this hypothesis, a detailed structural analysis of transporters is required. The structure of the P-glycoprotein has been determined by electron microscopy to a resolution of 25 Å [Rosenberg et al., 1997]. This resolution is too poor to describe a multidrug-binding site, yet one can discern several features of the structure that are consistent with biochemical and genetic data. From the extracellular face of the membrane, P-glycoprotein has a toroidal shape with a large aqueous pore, which is closed at the cytoplasmic face. The P-glycoprotein also forms an opening to the lipid phase within the plane of the membrane, which is consistent with the "flippase" model and the data showing that substrates can access the translocation pathway from the lipid bilayer phase. Nevertheless, this model doesn't explain how multidrug transporters can bind structurally unrelated compounds. This exceptional property of multidrug transporters can be proven directly only by structural analysis of a multidrug transporter in complex with its substrates. Hence, because of the dearth of structural information on any of these membrane proteins, the structural mechanisms of multidrug recognition remain obscure and require further investigation.

1.4. Regulation of multidrug resistance and multidrug-binding regulators.

Multidrug transporters are subject to complex regulatory systems. Despite scant knowledge of the regulatory pathways that affect *mdr* gene expression, it is known that in mammalian cells, selection for drug resistance usually results in overexpression of transporters, either by amplification of a wild-type gene, or alterations of gene expression that increase levels of wild-type *mdr* RNA [Shen et al., 1986], or both. However, in mammalian cells, the mechanism of regulation is poorly understood whereas more is known about the regulation of bacterial multidrug resistance. Namely, in prokaryotic cells, multidrug resistance can be induced by substrates of multidrug transporters.

In bacteria, several transcription regulators of multidrug transporters have been demonstrated to promote transporter expression in response to structurally dissimilar toxic compounds. For example, in the Gram-negative bacterium *Escherichia coli*, EmrR (**E**. coli **m**ultidrug **r**esistance **R**egulator) negatively regulates the transcription of the multidrug transporters EmrA and EmrB presumably through its direct binding of salicylates and nalidixic acid [Lomovskaya et al., 1995]. Similarly, the expression of the *E.coli* multiple-antibiotic resistance operon *marRAB* is repressed by MarR (**M**ultiple-**a**ntibiotic **r**esistance **R**egulator) but is restored by salicylate, dinitrophenol, and benzoate [Seoane & Levy, 1995]. A third efflux system from *E.coli*, AcrAB, is responsible for resistance to bile acids and fatty acids, such as decanoate. Expression of *acrAB* is induced by decanoate and is likely to be mediated by a transcription regulator AcrR [Ma et al., 1996]. In the Gram-positive bacterium *Staphylococcus aureus*, several substrates of the QacA multidrug transporter, including ethidium and benzalkonium cations, increase the expression of a *qacA* gene by a mechanism that is dependent on a transcription regulator, QacR, which binds these compounds and derepresses transcription of the *qacA* gene [Grkovic et al., 1998]. All of these multidrug-sensing systems ensure that an adequate number of transporters are available to decrease high intracellular concentrations of these toxic compounds.

Logically, these regulatory systems can function only if the transporter and the transcription regulator possess a multidrug binding site. Indeed, direct binding of multiple structurally dissimilar compounds has been demonstrated for two transcription regulators: QacR from *Staphylococcus aureus* [Grkovic et al., 1998] and BmrR from *Bacillus subtilis* [Ahmed et al., 1994]. Unlike multidrug transporters, which have yet to be purified in the amounts necessary for crystallographic studies, cytosolic transcription regulators would appear to be more amenable to structural studies and thus, provide promising model systems for understanding the structural principles underlying multidrug recognition. Although QacR is beyond the scope of this work, BmrR will be discussed here in detail.

2.1. Multidrug regulation in bacteria *Bacillus subtilis*.

As we now know, the phenomenon of multidrug recognition is not confined to multidrug transporters. In the Gram-positive bacterium *Bacillus subtilis*, the transcription regulator BmrR directly binds a number of structurally-dissimilar aromatic hydrophobic cations and activates transcription of the multidrug transporter gene *bmr* . Thus, B. *subtilis* utilizes a two-component system of multidrug-binding proteins, consisting of a transporter and its transcription regulator, to protect itself against multiple toxic compounds.

Besides Bmr, *B.subtilis* has another multidrug transporter, Blt. Bmr and Blt share 51% sequence identity and have very similar substrate spectra, but the patterns of expression for these two transporters differ [Ahmed et al., 1995]. Expression of Blt is controlled by transcription activators BltR [Ahmed et al., 1995] and MTA [Baranova et al., 1999]; however, their respective small molecule inducers are unknown. Blt, BltR, and MTA are beyond the scope of this work.

2.2. Identification of transcription activator BmrR.

BmrR was first identified as an open reading frame in the vicinity of the *bmr* gene. The product of this open reading frame encodes a protein the N-terminal region of which shows considerable homology to N-terminal DNA-binding regions of MerR-family proteins (described in detail below). Moreover, the *bmr* promoter contains an inverted repeat sequence and a 19bp spacer between its "-35" and "-10" elements (Figure 1.2) [Ahmed et al., 1994], hallmarks of DNA sites bound by MerR-family proteins [Amabile-Cuevas et al., 1991; Summers, 1992; Holmes et al , 1993]. Such an analogy with MerR suggested that the product of this open reading frame encoded a transcription regulator of

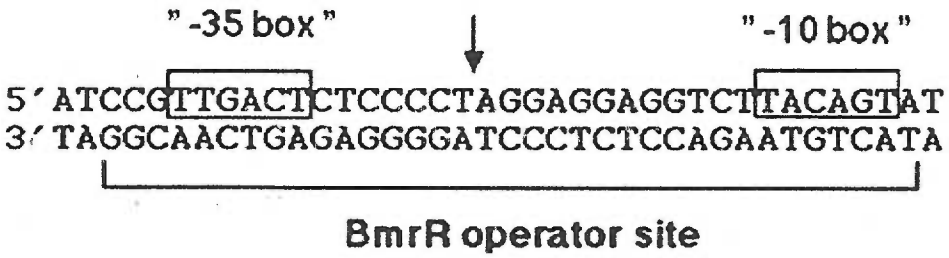


Figure 1.2. Promoter region of the *bmr* gene and BmrR binding site [Ahmed et al, 1994]. The arrow indicates a center of the pseudopalindromic operator.

bmr expression; hence, this putative regulator was named BmrR for a **Bmr R**egulator [Ahmed et al., 1994]. When BmrR was overexpressed and purified to homogeneity, in a series of footprinting and gel-shift assays, the purified BmrR indeed bound the previously identified inverted repeat located between "-35" and "-10" elements of the *bmr* promoter (Figure 1.2) [Ahmed et al., 1994].

2.3. Role of BmrR in *bmr* expression.

The critical role of BmrR in the expression of *bmr* gene has been indicated by the following experiment. When the *bmrR* gene is deleted or replaced with a chloramphenicol-acetyl-transferase (CAT) construct, the drug-resistant phenotype of wild-type cells is converted to a drug-sensitive phenotype; plus, no *bmr* mRNA can be detected in these cells [Ahmed et al., 1994]. Thus, BmrR is necessary for Bmr expression in *B.subtilis* cells. Furthermore, *bmr* expression can be induced by several Bmr substrates, e.g., rhodamine and tetraphenylphosphonium (TPP) (Figure 1.1) can stimulate transcription of the *bmr* gene by binding to BmrR [Ahmed et al., 1994]. More specifically, both rhodamine and TPP have been shown to induce transcription from the *bmr* promoter thereby increasing the levels of *bmr* mRNA by 20-fold in wild-type cells as well as to induce β-galactosidase expression in those cells where the *bmr* promoter controls the expression of the *lacZ* gene. Also, rhodamine and TPP increase by several fold *Bacillus subtilis* resistance to other substrates of Bmr, such as puromycin, norfloxacin, and ethidium bromide. Moreover, neither of the above results is observed in cells in which the *bmrR* gene has been disrupted [Ahmed et al., 1994]. Finally, rhodamine binds BmrR·promoter complex specifically with K_d of 1.5 μM while TPP competes for this binding with K_d of 100 μM [Ahmed et al., 1994]. Interestingly, neither rhodamine nor TPP change the BmrR binding pattern or its affinity for the cognate operator.

2.4. The MerR family of transcription activators.

The MerR family is primarily a "detoxification" family. The seven known proteins of the MerR family activate transcription of specific genes only when they bind a small molecule inducer, which is different for each protein. The prototype protein of the family, MerR, is found in several bacterial species where it activates the mercury resistance operon upon binding mercuric ion [Summers, 1992; Ansari et al., 1995; Utschig et al., 1995]. Another well-studied protein from this family, SoxR of *Escherichia coli*, senses the oxidative state of a cell through its 2Fe-2S clusters and turns on transcription of a regulon involved in repairing oxidative damage [Gaudu & Weiss, 1996 and Hidalgo et al., 1997]. In addition to the aforementioned BmrR, BltR, and MTA from *Bacillus subtilis*, which regulate multidrug resistance [Ahmed et al., 1994; Ahmed et al., 1995; Baranova et al., 1998], $TipA_L$ regulates thiostrepton resistance in *Streptomyces lividans*. $TipA_L$ stands apart from the rest of the family because its inducer thiostrepton binds covalently to the protein [Holmes et al., 1993]. Finally, the somewhat less studied NolA, from *Bradirhizobium japonicum*, activates expression of bean nodulation genes in the presence of several flavonoids [Sadowsky et al., 1991].

All members of the MerR family share significant homology at the N-terminal DNA-binding domain (Figure 1.3). These dimeric activator proteins use a putative helix-turn-helix motif to bind the inverted DNA repeats of the promoter regions of the genes that they regulate.

The MerR-type proteins are also delineated by the unusual charasteristics of their cognate DNA binding sites; the spacer region between the -35 (consensus sequence TTGACA) and the -10 (consensus sequence TATAAT) promoter elements is 19bp, which is longer than the common 16-18bp [Russell and Bennett, 1982; Gralla and Collado-Vides, 1996]. A promoter with such a long spacer region is inefficient because the -35 and -10

Figure 1.3. Amino acid sequence alignment of MerR family proteins. Black boxes designate identical and gray boxes homologous residues.

boxes are located on the opposite sides of the double DNA helix, and RNA polymerase cannot simultaneously contact them both. Interestingly, when the wild-type 19bp spacer is shortened by 2 base pairs, the resulting 17bp spacer yields an efficient promoter, functioning independently of a regulator. This has also been shown for promoters that are regulated by SoxR [Hidalgo and Demple, 1997] or MerR [Lund and Brown, 1989; Parkhill and Brown, 1990].

Combined studies on two members of the MerR family, MerR and SoxR, have led to the following model of transcription regulation (Figure 1.4). In the absence of an inducer, an inactive regulator binds between the -35 and -10 promoter elements of the regulated gene and allows the RNA polymerase to form an inactive or "closed" complex. In fact, Anzari et al has shown that inactive MerR bends DNA within the promoter, and that RNA polymerase contacts only the -35 element, thus forming a "closed" complex [1994]. However, upon binding a small molecule inducer, the regulator restructures its promoter by unbending and underwinding the DNA and positions the -35 and -10 elements on the same side of the DNA helix. In support, footprinting experiments with MerR have revealed that activated MerR distorts the suboptimal 19 bp spacer to make it more like a 17 bp spacer [Anzari et al., 1995], and chemical nuclease cleavage assays suggested that MerR underwinds the spacer [Anzari et al., 1992]. Only then can RNA polymerase form a transcriptionally active or "open" complex and proceed with transcription. This activator-dependent unbending and untwisting of the DNA remodels the promoter and makes it a better template for the poised polymerase. Although this model is very attractive for all MerR family members, it still has to be confirmed for BmrR.

Furthermore, the detailed structural mechanisms by which MerR-type proteins, including BmrR, bind their cognate promoters and small molecule inducers remains unknown.

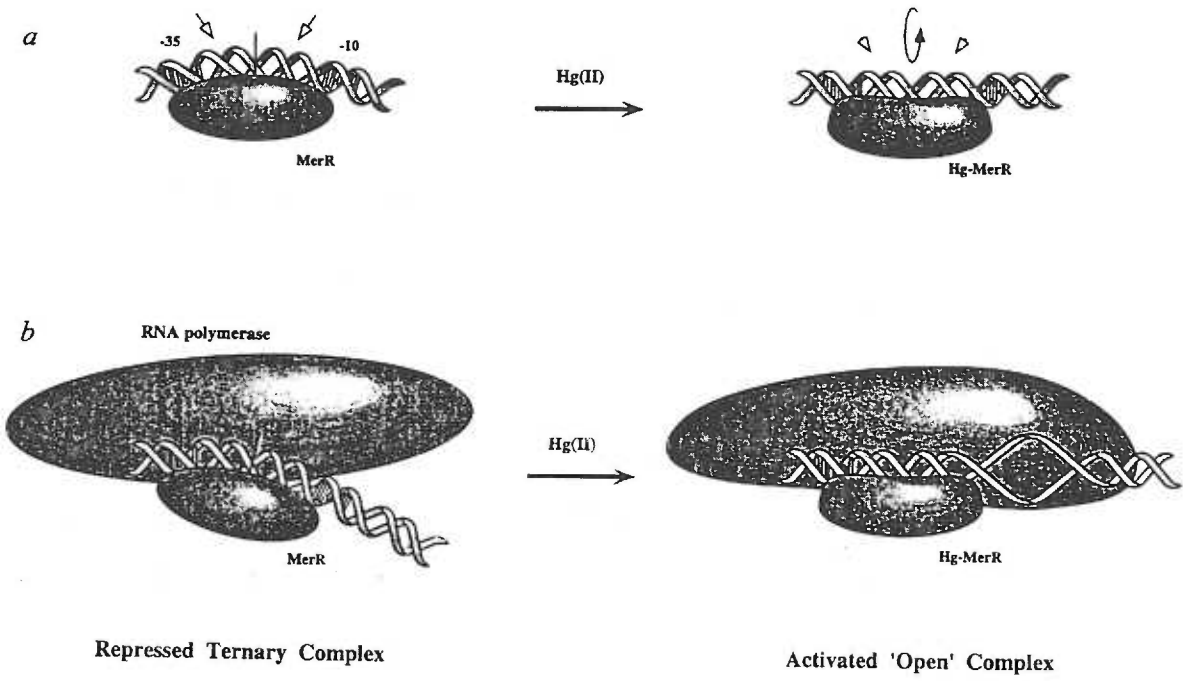


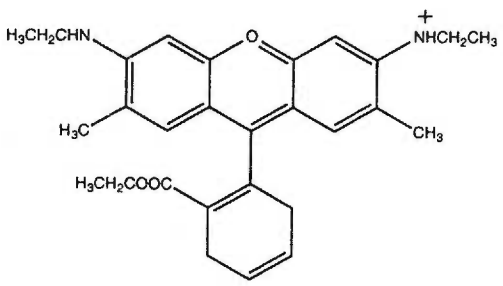
Figure 1.4. Model of transcription activation by MerR as proposed by Anzari et al, 1995 (adapted from Anzari et al, 1995).

2.5. Multidrug binding by BmrR and BRC.

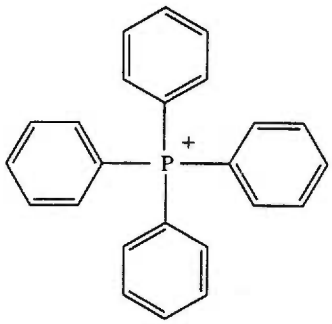
BmrR is a 279 amino acid protein that is composed of an amino-terminal DNA-binding domain (residues 1-90), a linker region (residues 91-120), and a carboxy-terminal multidrug-binding domain (residues 121-279). Whereas the carboxy-terminal activation domains of most MerR family proteins bind a single specific ligand, BmrR differs in its high-affinity binding of the structurally dissimilar molecules. As previously found, rhodamine binds purified BmrR-promoter complexes with a K_d of 1.5 μM. TPP can displace rhodamine from the BmrR-promoter complex and binds with K_d of 100 μM [Ahmed et al., 1994]. Consequently, rhodamine and TPP should bind to the same site on BmrR.

The C-terminal domain of BmrR (residues 121-279), named BRC (BmrR C-terminus), represents the multidrug-binding domain of BmrR, and it has no sequence homology to any other proteins in current databases. Overexpressed and purified BRC forms dimers and binds the aromatic cationic ligands rhodamine and TPP with the same affinity as the full-length BmrR [Markham et al., 1996].

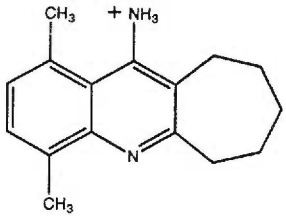
BRC has a very broad ligand specificity, which is comparable to the substrate specificity of multidrug transporters. In addition to rhodamine and TPP, BRC binds several other structurally dissimilar, cationic aromatic compounds (Figure 1.5). Among these are four potent inducers of BmrR identified through screening of 2,100 organic compounds in the DiverSet library [Markham et al., 1997]. Specifically, compounds designated I, II, III, and IV have been identified by their ability to induce growth of wild-type *B. subtilis* cells in the presence of ethidium bromide. All four compounds act analogously to rhodamine and TPP: they increase the ethidium-bromide resistance of wild-type *B.subtilis* cells and stimulate expression of β-galactosidase in the cells carrying *bmr* promoter-*lacZ* fusion construct. More importantly, these compounds bind BRC directly



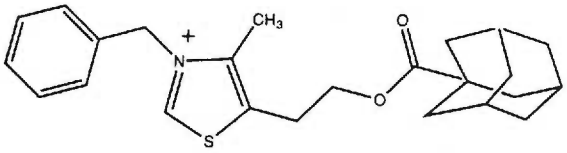
Rhodamine 6G



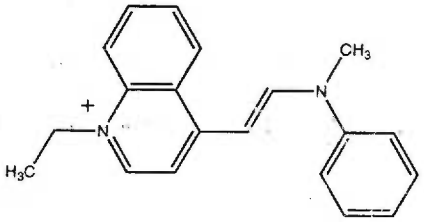
Tetraphenylphosphonium



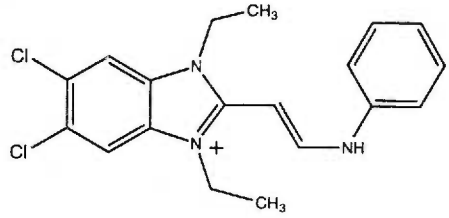
I.



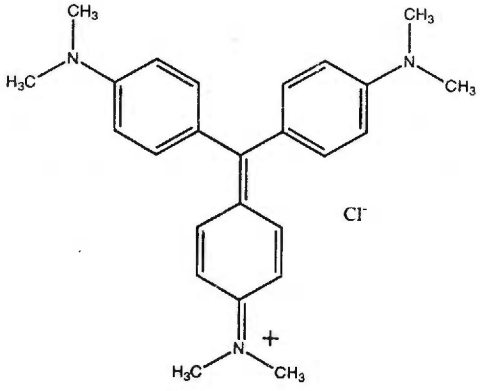
II.



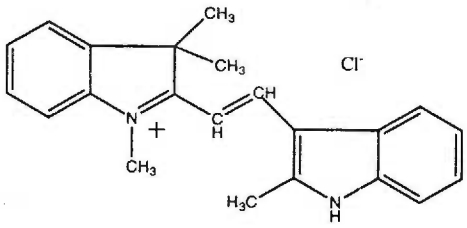
III.



IV.



Crystal violet



Astrazon orange

Figure 1.5. BRC ligands.

with high affinities: with K_d's of 1 μM (compound I), 55 μM (compound II), 50 μM (compound III), and 0.8 μM (compound IV). When these drugs are bound to the BRC domain, BmrR becomes a potent transcription activator.

A few compounds, which have been shown to bind BRC with high affinity, do not induce transcription [A. Neyfakh, personal communication]. Among these are crystal violet (K_d=0.8 μM) and astrazon orange (K_d=1.1 μM) (Figure 1.5). These latter compounds are interesting because they may provide insight into the role that drug-binding plays in the mechanism of transcription activation.

Thus, biochemical data indicated that BmrR existed as dimer and possessed an unusual multidrug-binding site. Moreover, the C-terminal domain (BRC) (residues 121-279) appeared to be solely responsible for the binding of different drugs, and it could form dimers. Since ligands competed for binding to BRC and since the stochiometry of binding was 1:1 (drug per monomer), we hypothesized that BmrR has a single multidrug-binding site located on its C-terminus, within the BRC domain. However, nothing was known either about the location of the binding site or about the mechanism by which structurally unrelated compounds interact with this multidrug-binding site. To reveal the nature of the intriguing multidrug-binding site in atomic detail, we used x-ray crystallography. The multidrug-binding domain BRC is an excellent model for the multidrug recognition by the full-length BmrR, because it functions autonomously from the DNA-binding domain and because it is more soluble and thus, amenable for crystallization. In the first part of the following dissertation, I have focused primarily on determining the structures of apo BRC and BRC-drug complexes by x-ray crystallography.

B. **DNA-binding proteins.**

The specificity of protein-DNA interactions is critical in regulating cellular processes involving DNA transcription, replication, repair, and recombination. In the crowded interior of a cell packed with proteins working on their own particular tasks, every protein must interact only with a limited number of appropriate molecules and not with any other even if the latter are present in high concentrations. Each of these proteins is a precise machine working on a subnanometer scale. To be able to understand and, perhaps, influence cellular events, we need to know the rules governing the specificity of protein-DNA interactions.

1.1. Specific protein-DNA contacts.

Depending on requirements imposed by their functions, DNA-binding proteins select their cognate target DNA sequences with specificities ranging from stringent to fairly permissive, which is reflected in their affinities varying from 10^{-8} M to 10^{-12} M. How do proteins achieve specificity? In other words, how can they distinguish the correct and incorrect sequences apart? Structural and biochemical studies have shown that the reading of a DNA sequence by a protein mostly depends on 3 interdependent factors: sequence-specific direct interactions between protein and DNA, shape complementarity, and an indirect DNA effect [Seeman et al., 1976; Rhodes and Burley, 1997]. The specific interactions include hydrogen bonds and some van der Waals interactions. Hydrogen bonds are formed between polar functional groups of the DNA bases (such as O6 group of guanine) and polar protein side chains (such as arginine) [Luisi et al., 1991; Schumacher et al., 1994] or main chains [Mandel-Gutfreund et al., 1995]; however, recently, CH-O hydrogen bonds have also been implemented in protein-DNA recognition, e.g. hydrogen bonds between the 5-methyl group of thymine and a carbonyl oxygen of a protein [Mandel-Gutfreund et al., 1998]. In some cases, hydrogen bonds mediated by water can also

account for the specificity [Otwinowski et al., 1988; Shakked et al., 1994]. In turn, specific van der Waals interactions can be formed between nonpolar groups on the base (such as the 5-methyl group of thymine) and a nonpolar protein side chain (such as a valine side chain) [Luisi et al., 1991] although sometimes, van der Waals interactions are observed between nonpolar and polar groups.

1.2. Surface complementarity.

DNA-binding proteins exhibit remarkably diverse protein architectures, yet all achieve a precise complementarity of shape. Tight shape complementarity of protein and DNA surfaces in a complex is advantageous because it maximizes the number of specific protein-DNA contacts, thus facilitating specific chemical recognition of a DNA target by a protein. In addition, shape complementarity maximizes the number of protein-DNA backbone contacts and increases the buried surface area of the protein-DNA complex. When protein and DNA form a complex, part of the hydrophobic surface area becomes protected from the solvent buried at the interface. The resulting entropically favourable hydrophobic effect improves the stability of specific complexes and in this manner affects specificity [Calladine and Drew, 1992; Spolar and Record, 1994].

To optimize the fit of their recognition surfaces, both protein and DNA undergo conformational adjustments in a manner called "induced fit" [Koshland, 1962; Frankel and Kim, 1991]. Conformational changes, observed in proteins, include ordering of the disordered loops and formation of α-helices from unfolded regions [Ferre-D'Amare et al., 1994; Schumacher et al., 1994; Spolar and Record, 1994; Spronk et al., 1996]. Interestingly, in protein-DNA complexes, the DNA molecules sustain multiple kinds of distortion such as bending, kinking, unstacking, underwinding, and overwinding [Schultz et al., 1991; Kim et al., 1993a; Kim et al., 1993b; Rice et al., 1996].

1.3. DNA bending by proteins.

Many DNA-binding proteins manipulate DNA structure to gain better access to the DNA bases, to optimize favourable interactions between the protein and the DNA, and to increase the buried surface area, all of which are necessary for specificity and affinity of protein-DNA interactions. Thus, some distortion of the DNA structure is inherent to the binding of most proteins. However, only some proteins distort DNA to an extreme, and frequently, such severe DNA distortion is an integral part of their function: transcription factors, enzymes that act upon DNA, and DNA-packaging proteins.

Transcription factors regulates transcription by bending and looping DNA in various ways: they bring distantly-bound transcription factors into proximity with RNA polymerase and with each other or they facilitate additional interactions between DNA and the general transcription machinery. For example, the prokaryotic transcription activator CAP (catabolic activator protein) bends DNA by a total of 90° through 45° bends at two CpA steps [Schultz et al., 1991]. As another notable example, a eukaryotic transcription factor TBP (TATA-binding protein) unwinds and rolls open the minor groove of its target sequence resulting in an 80° bend away from the protein [Kim et al., 1993a; Kim et al., 1993b]. TBP completely unstacks the base steps at two symmetry-related kinks and inserts phenylalanine side chains at each kink. Architectural proteins, e.g., the integration host factor IHF, manipulate DNA in a manner similar to transcription factors: they loop DNA thus facilitating protein-protein interactions in many cellular processes including transcription initiation. IHF wraps the DNA around itself and bends it by over 160 degrees. The bending occurs at two large kinks where the base stacking is interrupted by intercalation of a proline residue [Rice et al., 1996].

Many restriction endonucleases significantly deform their substrate DNA. As an example, EcoRI kinks DNA in the middle of its palindromic DNA recognition site (ApT step) [Kim et al., 1990]. Such kinking opens the major groove for direct protein-base contacts and orients the phosphate backbone for a hydrolytic attack. Another restriction

endonuclease, EcoRV, acts somewhat differently, sharply bending the DNA and compressing the major groove, yet it achieves the same goal: to bring the phosphate group into the vicinity of the catalytic site [Winkler et al., 1993].

In addition, proteins may wrap DNA for a topological purpose. Many specific processes, such as initiation of transcription, recombination, and replication require the formation of nucleoprotein complexes in which the DNA is tightly wrapped. In the nucleosome core particle, the histone octamer dramatically bends DNA at multiple sites to wrap 146 base-pairs of DNA in 1.8 superhelical turns [Luger et al., 1997]. Other proteins may kink DNA only to achieve surface complementarity. PurR and LacI exemplify these proteins. The question which remains is, if the DNA gets distorted in the protein-DNA complex, what are the properties of the DNA that allow such distortion and also influence the specificity?

Traditionally, only surface complementarity and direct protein-DNA base interactions have been considered as contributing to the specificity. However, in recent years, a number of studies [Rice et al., 1996; Bailly et al., 1996] have suggested and discussed the importance of indirect influence of DNA on protein specificity. For example, bases that are not contacted by a DNA-binding protein may affect the affinity of protein for the DNA by changing structural parameters of the DNA, as observed in EcoRI [Lesser et al., 1990], 434 repressor [Chen et al., 1997], and LacI [Zhang and Gottlieb, 1995]. Sequence dependent deformability plays a role in the specificity of the integration host factor IHF for its cognate DNA, as evidenced by the dearth of direct protein-DNA base contacts [Rice et al., 1996].

1.4. *E.coli* purine repressor, PurR.

The *E.coli* purine repressor (PurR) is the principal regulator of *de novo* purine nucleotide biosynthesis and, to a lesser extent, *de novo* pyrimidine nucleotide biosynthesis [Zalkin and Dixon, 1992; Zalkin and Nygaard, 1996] (Figure 1.6). Specifically,

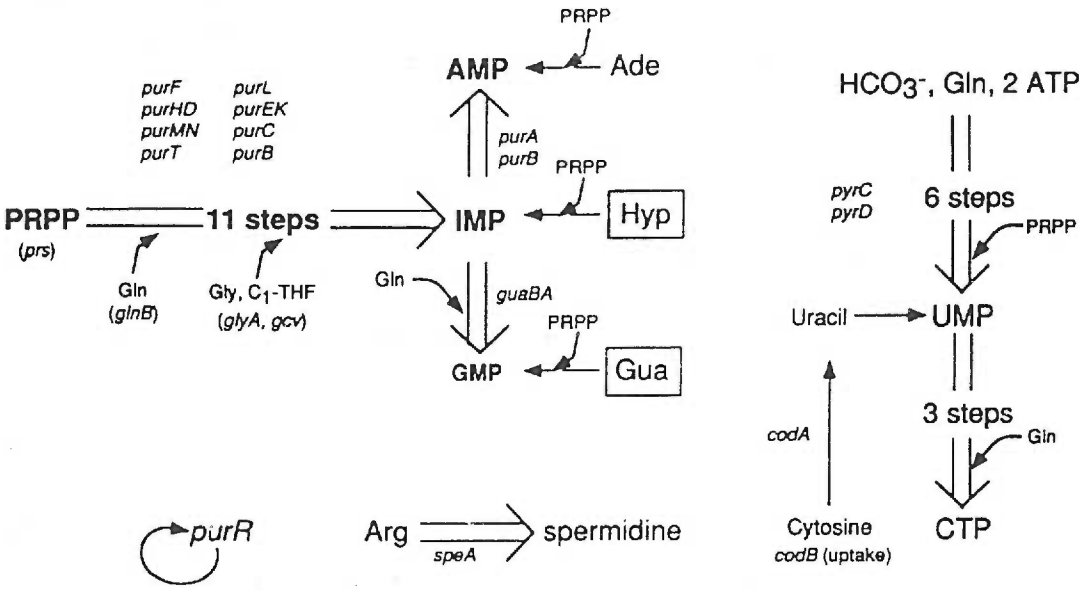


Figure 1.6. PurR regulatory circuit (adapted from Zalkin and Nygaard, 1996).
Hyp stands for hypoxanthine, Gua – guanine, Ade – adenine, PRPP – phosphoribosyl pyrophosphate, AMP – adenosine monophosphate, GMP – guanosine monophosphate, IMP – inosine monophosphate, UMP – uridine monophosphate, CTP – cytidine triphosphate.

PurR represses transcription from ten *pur* operons that encode enzymes of purine biosynthesis [He et al., 1990; Zalkin and Nygaard, 1996] and several enzymes of pyrimidine biosynthesis and salvage [Choi and Zalkin, 1990; Wilson and Turnbough, 1990] as well as several other genes encoding proteins necessary for polyamine and purine nucleotide biosynthesis [He et al., 1993]. In addition, PurR is autoregulated [Rolfes and Zalkin, 1990].

PurR represses transcription by binding to a pseudopalindromic operator located upstream of the transcription start site. To bind its 16 base-pair operator [Rolfes and Zalkin, 1988] with high affinity, PurR must first bind one of its two corepressors guanine or hypoxanthine (Kd=1.5 μM and 9.3 μM, respectively) [Choi and Zalkin, 1992]. PurR, which exist as a 76 kDa dimer, binds to the *purF* operator, which controls the expression of the first enzyme in purine biosynthesis, glutamine-PRPP-amidotransferase, with the highest affinity. The Kd of this interaction is 2.5 nM as determined by fluorescent anisotropy [Glasfeld et al., 1996].

PurR belongs to LacI family of bacterial repressors [Weickert and Adhya, 1992] and can be divided into two functional domains: the N-terminal DNA-binding domain (residues 1-60) and the C-terminal corepressor-binding domain (residues 61 to 340) [Schumacher et al., 1994] (Figure 1.7). The corepressor-binding domain (CBD) is structurally similar to bacterial periplasmic-binding proteins (PBP) [Spurlino et al., 1991; Oh et al., 1993; Schumacher et al., 1993] and especially to ribose-binding protein RBP [Mowbray and Cole, 1992; Bjorkman and Mowbray, 1998].

The CBD is solely responsible for the binding of corepressors. As seen in periplasmic-binding proteins, PurR binds corepressors in a cleft between 2 structurally similar N- and C-terminal CBD subdomains through a series of polar, nonpolar, and aromatic contacts (Figure 1.8, top). Our lab has solved the structure of PurR bound to the *purF* operator and hypoxanthine [Schumacher et al., 1994] and guanine [Schumacher et

Figure 1.7. Structure of the PurR-hypoxanthine-*purF* operator complex [Schumacher et al., 1994]. Labelled are the corepressor-binding domain (CBD) and the DNA-binding domain (DBD). The CBD is coloured red, the helix-turn-helix-loop-helix motif is coloured yellow, and the hinge helix is coloured green. The hypoxanthine corepressor is shown as cyan balls and sticks.

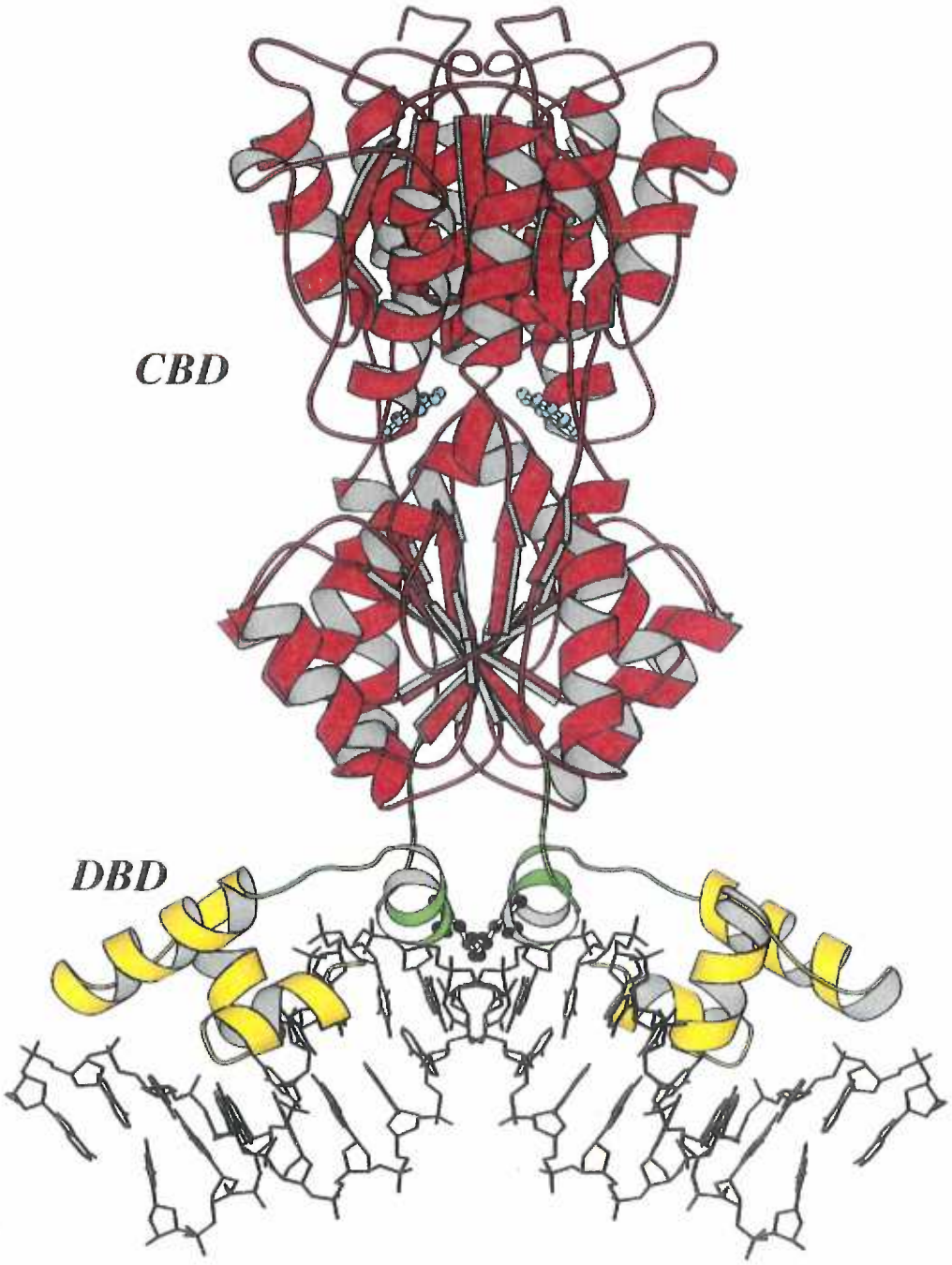
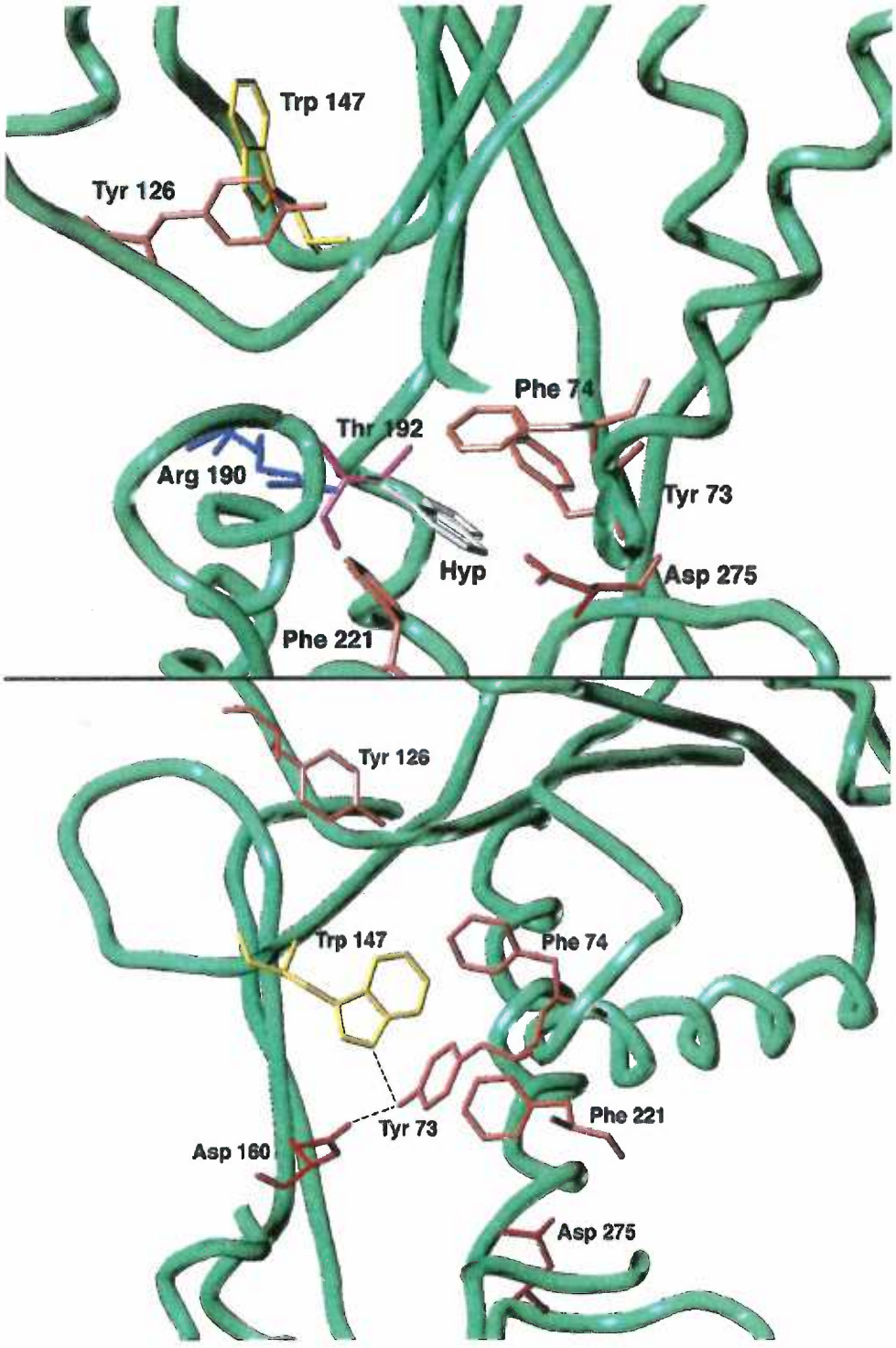

CBD
DBD

Figure 1.8. The corepressor-binding pocket of the corepressor-free and the corepressor-bound PurR. The Cα trace of PurR is shown as a green worm. Trp147 residue is shown in yellow, hypoxanthine molecule (top) is shown in white.


Trp 147
Tyr 126
Phe 74
Thr 192
Arg 190
Tyr 73
Hyp
Asp 275
Phe 221
Tyr 126
Trp 147
Phe 74
Tyr 73
Phe 221
Asp 160
Asp 275

the presence of DNA. Thus, PurR features intramolecular signalling between the distant corepressor-binding and DNA-binding domains.

PurR binds its pseudopalindromic operator as a dimer where the 60-residue N-terminal DNA-binding domain of each monomer contacts an operator half-site [Schumacher et al., 1994]. Actually, the DNA-binding domain consists of 2 DNA-binding elements: the helix-turn-helix (HTH) motif, which binds in the major groove, and the hinge helix, which binds in the minor groove (Figure 1.9).

Residues of the helix-turn-helix make a large number of direct and water-mediated contacts in the major groove of the purF operator. These DNA-base/protein interactions include residues Thr15, Thr16, and His20, from the HTH motif, and Arg26, located on the loop that follows the HTH motif. This last residue selects guanine at the position 4 in all binding sites of PurR. This HTH motif together with helix 3 exists as a stable globular domain even in the corepressor-free PurR, and possibly initiates specific DNA binding by making appropriate base-pair contacts with the operator [Nagadoi et al., 1995].

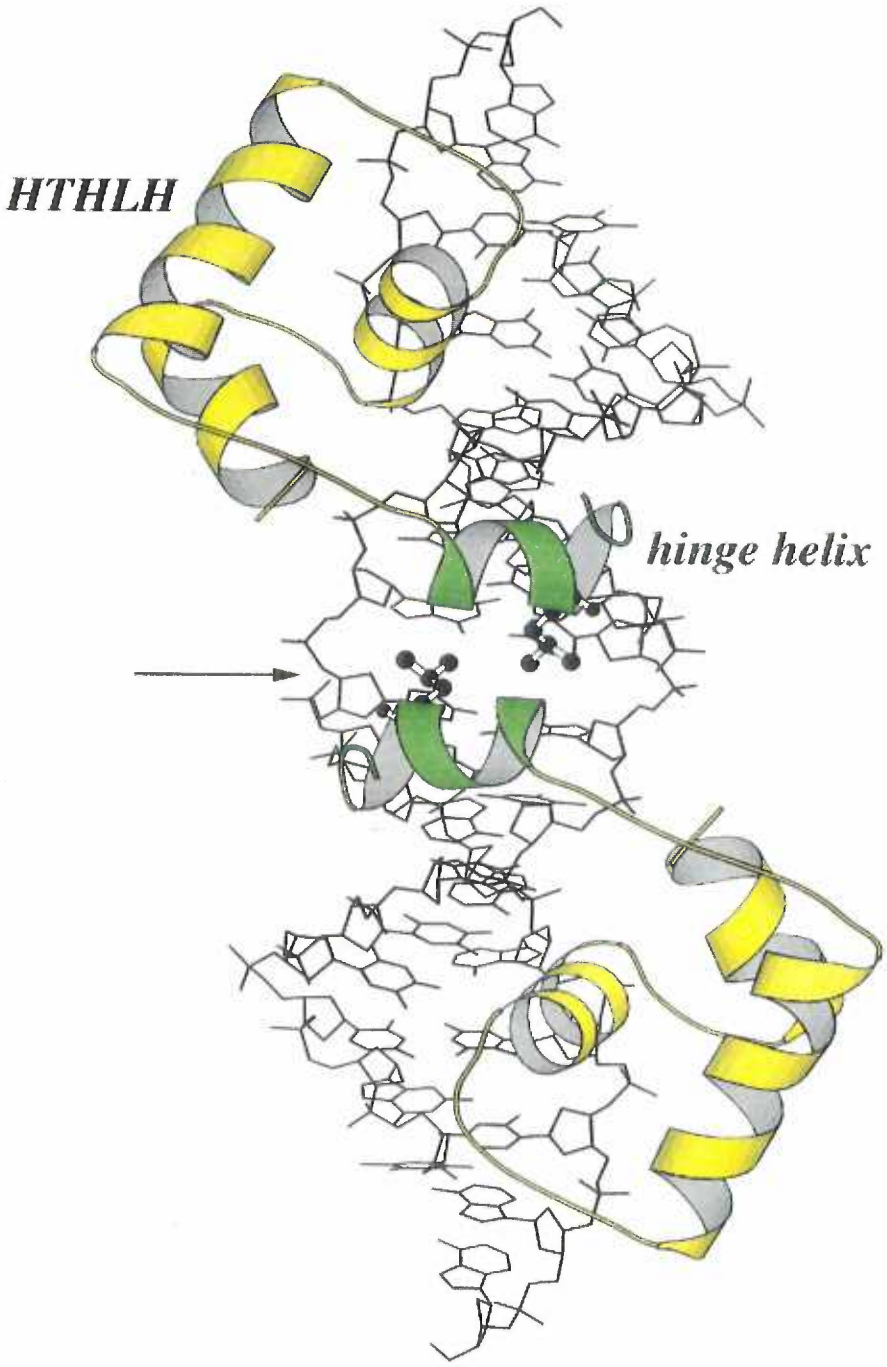
Residues of the hinge helix (residues 48-56) bind in the minor groove of the *purF* operator. In the PurR dimer, two symmetric hinge helices wedge into the minor groove at the center of the pseudopalindromic binding site with the following sequence (central step is shown in bold):

5'-A2 A3 G4 A5 A6 A7 A8 **C9 • G9'**T8'T7'T6'G5'C4'G3'T2'-3'

3'-T2'T3'C4'T5'T6'T7'T8'**G9'• C9** A8 A7 A6 C5 G4 C3 A2 -5'.

Two hinge helices (one from each monomer) intercalate the side chains of symmetry-related Leu54 residues into the central $Cyt^{9}pGua^{9'}$(CpG) base pair step of the *purF* operator (Figure 1.9). Because the hinge helices are too large to fit into the 5.7 Å-wide minor groove of B-DNA, they expand the minor groove and significantly distort the purF operator. Thus, two central C·G base pairs are unstacked, undertwisted, and rolled out to 48°; overall, the operator DNA is sharply kinked away from the protein. In addition

Figure 1.9. View of the two-fold related hinge helices and the minor groove of the *purF* operator. The arrow points at the location of the central CpG step. Hinge helices are shown in green, and helix-turn-helix-loop-helix (HTHLH) domains are shown in yellow. DNA is shown in black.


HTHLH
hinge helix

to the Leu54 interaction with the central C·G base, hinge helix residue K55 of PurR is key to minor groove binding specificity: it discriminates against guanine in favour of adenine at position 8 of the *purF* operator. Specifically, K55 reaches into the minor groove and makes a weak hydrogen bond and a vdW contact with N3 and C2 of Ade8, respectively. Accordingly, the mutation K55A decreases the binding affinity because it removes the positive charge from the PurR even though it doesn't change the overall architecture of the PurR-DNA complex [Glasfeld et al., 1996].

Proteolysis studies [Choi and Zalkin, 1994] have revealed that the hinge helix region is highly susceptible to cleavage only when PurR in not bound to DNA (specific binding) – a strong indication that DNA is a prerequisite for the stable formation of hinge helices. Furthermore, recent NMR studies on the DNA-binding domain of PurR have demonstrated that in the absence of DNA, hinge residues form no discernable secondary structure. Such coil-to-helix transition in PurR upon binding its operator DNA is an example of induced fit mechanism, where both protein and DNA undergo conformational changes to create tightly complementary interface and stabilize the protein-DNA complex.

The central CpG step is conserved among all PurR operators [He et al., 1990]; moreover, it is conserved among the operators of nearly all LacI family members [Rolfes and Zalkin, 1988]. In fact, replacing the CpG step in the *purF* operator with a TpA step results in a complete loss of repression *in vivo* . Yet, in the crystal structure, PurR makes no "sequence-specific" contacts to CpG. Instead, hinge-helix residue Leu54 makes van-der-Waals contacts to the O2 atom of Cyt^{9} (3.4 Å), to the deoxyribose O4' atom of $Gua^{9'}$ (3.3 Å), to the N3 atom of $Gua^{9'}$ (3.9 Å), and to the N2 group of $Gua^{9'}$ (4.0 Å). Of these four contacts, only the last contact cannot be made in a TpA step. However, this single van der Waals contact alone is unlikely to account for the high specificity of PurR for the central CpG step and its discrimination against TpA.

1.5. Modified oligonucleotides in studies of specificity.

To understand in more details the thermodynamics of protein-DNA binding and the specificity of a given protein for a DNA sequence, it is necessary to estimate the role of specific moieties on bases that enable the protein to discriminate between binding to a "correct" and "incorrect" sequences. Using synthetic chemistry, one can make precisely defined modifications in certain positions on the DNA and then measure their effect on binding characteristics [Jen-Jacobson, 1995]. This method takes advantage of introducing single-substituted analogues into the DNA, which are structurally isosteric and should distort the DNA only minimally.

A single-base analogue can be used to delete a functional group, e.g., a 2-amino group of guanine in the minor groove (guanine->hypoxanthine) or a 6-amino group of guanine in the major groove (guanine->2-aminopurine). Alternatively, one can add a functional group to a base (adenine ->2,6 diaminopurine or cytosine->5-methylcytosine). The first modification introduces a Watson-Crick hydrogen bond whereas the second modification introduces the bulkier methyl group to where only hydrogen was previously found. Such deletions and additions can be used to probe potential sites of direct protein contact to bases, but these types of modifications also can affect Watson-Crick hydrogen bonds and the DNA conformation, i.e. the indirect readout properties of a DNA sequence.

Base analogues can be used to reverse a hydrogen-bonding polarity, that is, to introduce a hydrogen bond donor in place of an acceptor. If this approach is used, one must decide what to do with a complementary base: e.g. G->DAP modification reverses the hydrogen-bond polarity at position 6 of the purine ring. If a complementary C is retained, one Watson-Crick hydrogen bond is lost. However, a substitution of U or T for C will maintain the three hydrogen bonds.

In addition to the aforementioned modifications, one can modify a functional group by, for example, alkylation or introduce incorrect or mismatched base pairs. Finally, one can estimate the contribution of individual DNA groups in binding of a protein by

comparing protein binding to its natural recognition site versus modified sites. As an example, one can measure equilibrium binding of protein to its natural recognition site and modified DNA sites and calculate the $\Delta\Delta G_{bind}$.

The use of synthetic base analogues has allowed a number of investigators to identify the roles of exocyclic groups in DNA recognition by proteins such as restriction endonuclease EcoRI [Lesser et al., 1990], tryptophan repressor TrpR [Mazarelli et al., 1992], and lac repressor LacI [Zhang and Gottlieb, 1995]. Besides studying protein-DNA interactions, base analogues have been successfully used to probe the role of the exocyclic groups in the specificity of drug binding [Bailly et al., 1995; Bailly et al., 1998] as well as in intrinsic properties of DNA, such as minor-groove width [Bailly et al., 1995].

In conclusion, the *E. coli* purine repressor PurR is our model system for studying the thermodynamic characteristics and specificity of DNA-binding by proteins. Despite our significant biochemical and structural knowledge of this system, we do not know how PurR can recognize the CpG sequence as the correct one because it makes no specific contacts to that sequence. We hypothesized that since the chemical composition of the CpG and TpA sequences has distinct differences, specific groups at certain exocyclic positions affect PurR binding by an unknown mechanism. Therefore, my project aimed to identify specific exocyclic groups in both CpG and TpA sequences that affect PurR binding and to determine the likely mechanism through which they dictate the preference for the CpG sequence.

CHAPTER 2.

MACROMOLECULAR CRYSTALLOGRAPHY.

Crystals have fascinated people since ancient times, and because of the mystery surrounding them, crystals have even been attributed supernatural properties. However, early in the 20th century, one of their most remarkable properties - diffraction of x-rays - was discovered. This discovery founded crystallography, a powerful method of investigating the internal structure of crystals and of their constituent molecules in atomic detail, and thus, it has revolutionized the way we think of molecules. In particular, macromolecular crystallography has brought biochemistry to a new level of sophistication by showing exactly what proteins and nucleic acids looked like.

1. Basic definitions and concepts.

Crystals.

A glance at any crystal quickly reveals the symmetric arrangement of its sharp faces and edges. This external crystal symmetry originates from the internal periodic arrangement of molecules. In their crystalline state, unlike in solution, molecules take regular positions. Hence, a crystal is an ordered three-dimensional array of molecules held together by noncovalent interactions.

The smallest and the simplest volume element in a crystal is the unit cell, the elementary building block of a crystal. One can think of a crystal as an arrangement of many identical unit cells stacked together in three directions.

Depending on the symmetry of the internal molecular arrangement, a crystal can belong to one of seven crystal systems: triclinic, monoclinic, orthorhombic, tetragonal, cubic, hexagonal, or trigonal. By convention, a unit cell of a crystal is described by six

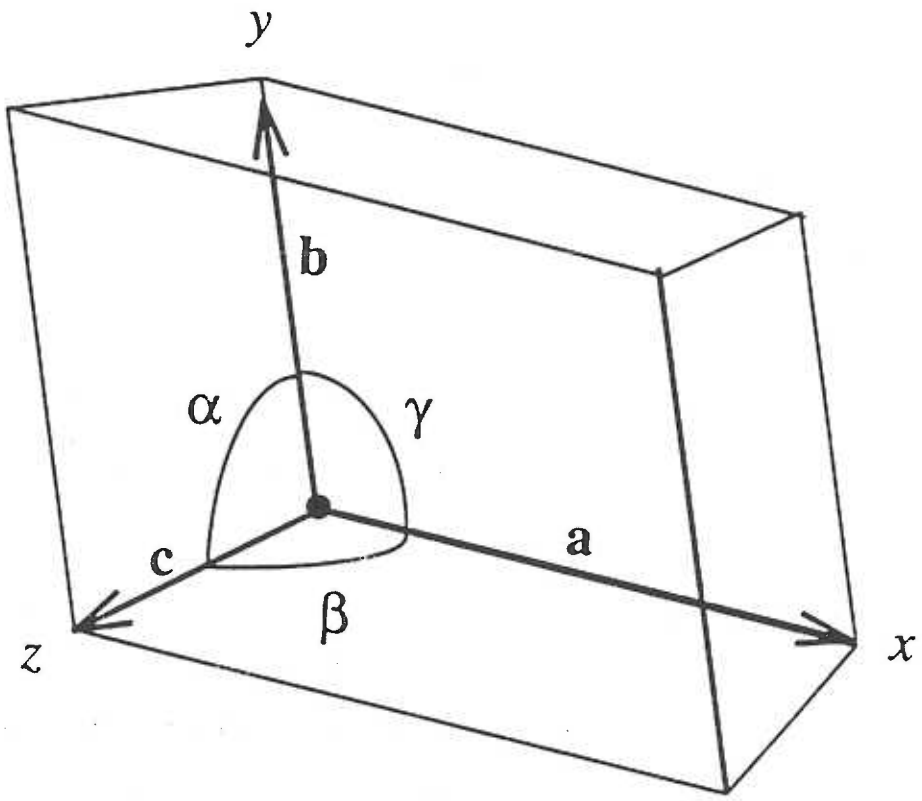


Figure 2.1. General unit cell with edges a, b, and c and angles α, β, and γ.

parameters: the lengths of three cell edges a, b, and c; and three cell angles α, β, and γ (Figure 2.1). A cell in which all six parameters can assume any value ($a \neq b \neq c$ and $\alpha \neq \beta \neq \gamma$) is called triclinic. If $a \neq b \neq c$, $\alpha = \gamma = 90^\circ$ and $\beta > 90^\circ$, the cell is monoclinic. A cell with $\alpha = \beta = \gamma = 90^\circ$ can be orthorhombic if $a \neq b \neq c$, tetragonal if $a = b \neq c$, or cubic if $a = b = c$. A hexagonal cell is characterized by $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$, and $a = b \neq c$. Last, a trigonal cell either has $a = b \neq c$ and $\alpha = \beta = \gamma \neq 90^\circ$ or has $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$, and $a = b \neq c$.

The array of points at the corners of unit cells is called the crystal lattice. Such a lattice is called primitive (P) if it has lattice points only at the corners of a unit cell. Each lattice point at the eight corners is shared with its eight nearest neighbours, which is equivalent to one point per unit cell. A simple primitive lattice is compatible with each of the seven crystal systems described above. However, some lattices can have more than one lattice point per unit cell, which makes them nonprimitive. If a lattice has additional lattice points on two opposing faces of the unit cell, it is called face-centered and is designated A, B, or C depending on which face is centered. In addition, a lattice can have all its faces centered (F) or can have its body centered (I). Nonprimitive lattices also conform to the symmetry of the seven crystal systems. In combination, seven unique nonprimitive lattices together with seven primitive lattices constitute 14 Bravais lattices, which are used to describe the internal symmetry of a crystal (Figure 2.2).

Any crystal can also be classified in terms of the group of symmetry operations (which include rotation and reflection) relating their faces. Each of these groups, known as point groups, represents one of the possible unique combinations of crystallographic symmetry elements. There are 32 point groups, and they are divided among the seven crystal systems.

In the late 19th century, mathematicians showed that the combination of 32 point

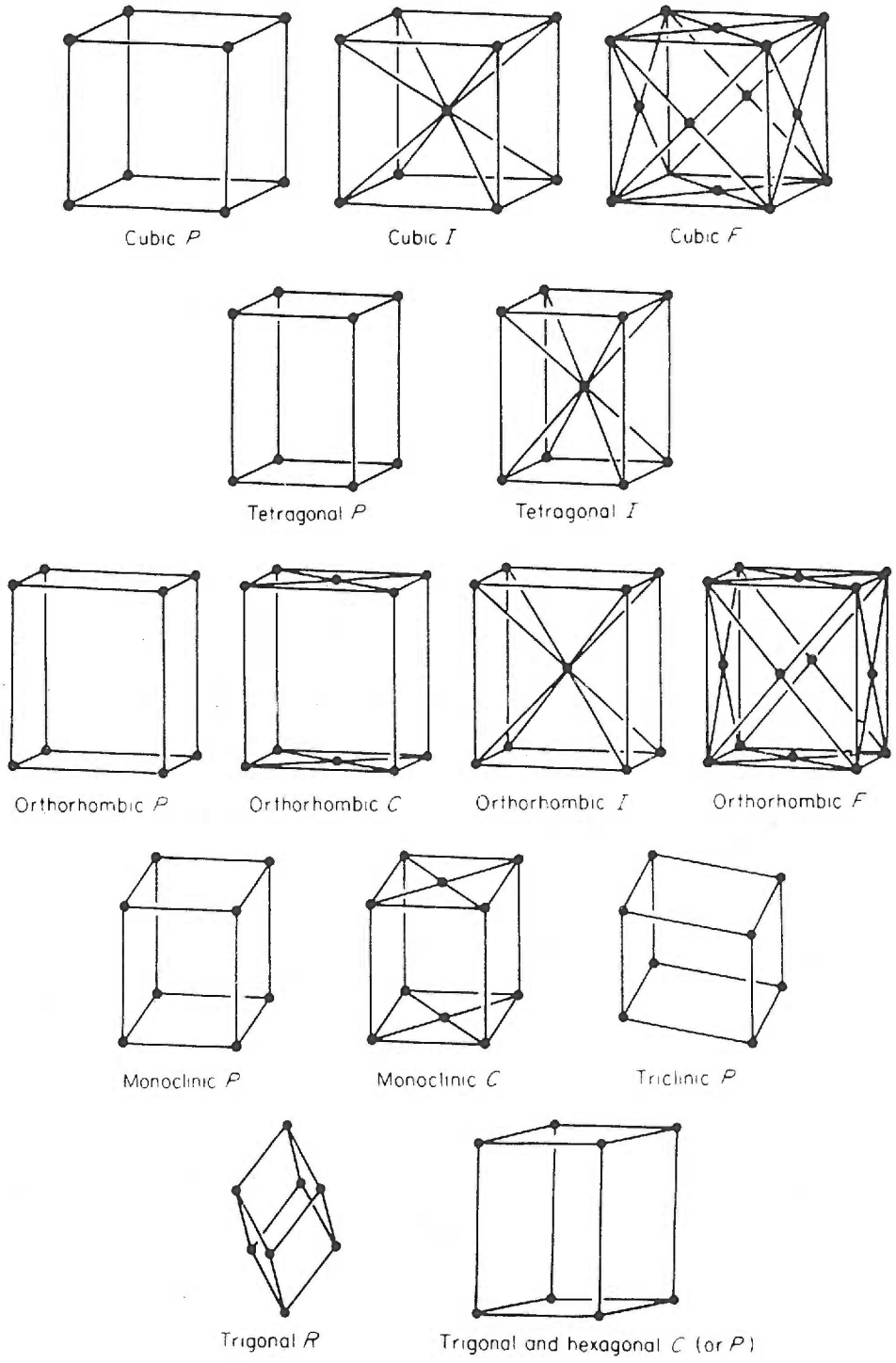


Figure 2.2. The 14 Bravais lattices (adapted from Blundell and Johnson, 1976).

groups with 14 Bravais lattices leads to 230 unique space groups, which describe the only possible ways to arrange identical objects in an infinite lattice such as crystal. These 230 space groups result from a combination of point groups with translational elements, which include screws and glide planes. However, in proteins, certain symmetry elements such as mirrors and inversion centers cannot occur because amino acids are asymmetric and exist in proteins only in their L-form. This limits the 230 possible space groups to 65.

A crystal space group is represented by a cryptic symbol (e.g., $P4_12_12$), in which a capital letter indicates a lattice type (P-primitive) and the other symbols represent symmetry operations that can be carried out within the unit cell. If the crystal contains symmetry elements, the unit cell will contain more than one asymmetric unit. For example, in space group $P4_12_12$, there are eight asymmetric units. An asymmetric unit of a space group is the basic part of a unit cell; one can create the whole unit cell by applying all symmetry operators of the space group to that asymmetric unit. Consequently, an asymmetric unit contains all the information necessary for the complete description of the crystal structure. This means that a crystallographer has to determine the contents of only one asymmetric unit in a crystal and then can reproduce the contents of the whole unit cell simply through symmetry operations.

An asymmetric unit can be occupied by one or more molecules. In this dissertation, both crystals of apo BRC and BRC-TPP crystals contained only one protein molecule in their respective asymmetric units. However, a noncrystallographic symmetry axis (e.g., relating monomers within a biological dimer or relating two halves of a DNA palindrome) can sometimes coincide with a crystallographic symmetry axis. In this case, such a symmetric molecule is in a "special" position, and the asymmetric unit will contain only a half of the molecule. For example, in the crystals of dimeric PurR bound to its cognate DNA pseudopalindrome, the 2-fold crystallographic symmetry axis passes through the PurR-DNA complex; in this case, the asymmetric unit cell is occupied by a PurR monomer bound to one-half of the DNA pseudopalindrome.

Why is it important to determine the symmetry of a crystal? In any array of lattice points, it is possible to choose a primitive triclinic cell regardless of the symmetry. However, to ignore the symmetry would be shortsighted. It is important to know the symmetry of a crystal because this knowledge simplifies the processes of interpreting the diffraction pattern, designing the strategy for the data collection, and solving the crystal structure. The importance of symmetry will become clearer below.

Diffraction.

For an object to diffract light, the wavelength of the light must be no larger than the object. Visible light has a wavelength of 400-700 nm and is not appropriate for the diffraction of atoms in molecules, which are about 0.15 nm apart. Wavelengths of x-rays are in the range 0.01-10 nm, and thus, x-rays can be diffracted by even the smallest molecules. However, it is impossible to produce a focused image of a single molecule because X-rays, unlike visible light, cannot be focused by lenses; plus, a single molecule is a very weak diffractor of x-rays. The problem is solved by analyzing the diffraction of crystals, which contain many copies of each molecule. Thus, the cumulative diffraction of a crystal is stronger than the diffraction of an individual molecule.

X-ray diffraction by crystals was discovered in 1912 by Max von Laue. In the same year, W. Laurence Bragg noticed the similarity of diffraction to reflection and deduced an equation regarding diffraction as "reflection" from parallel planes in the crystal lattice [1913]. (Even today, a spot on a diffraction pattern is called a reflection.) In Bragg's honour, this equation was later named Bragg's law.

The geometric construction (Figure 2.3) allows one to derive Bragg's law. If x-rays strike the crystal planes at an angle θ, they will be reflected also at an angle θ (see figure). The rays reflected from successive planes in the crystal will be in phase with each other and will produce a strong beam only if the path difference between the diffracted x-rays is equal to an integral number of wavelengths of the incident x-rays. This

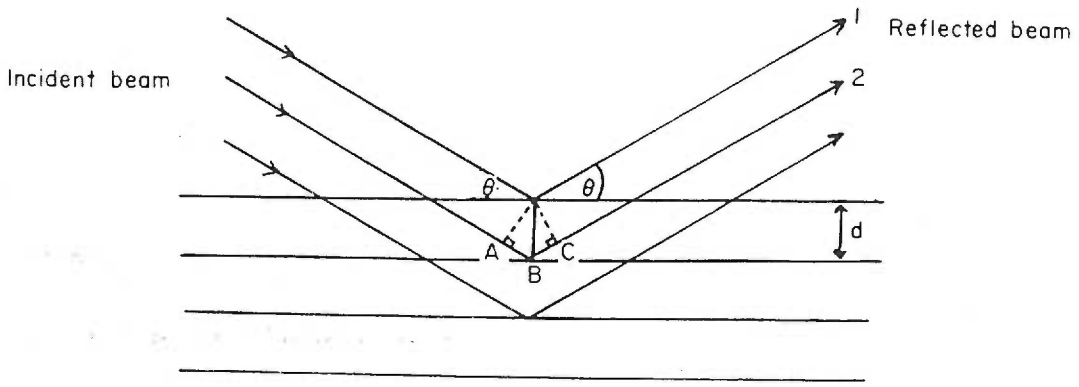


Figure 2.3. Bragg's law (adapted from Blundell and Johnson, 1976).

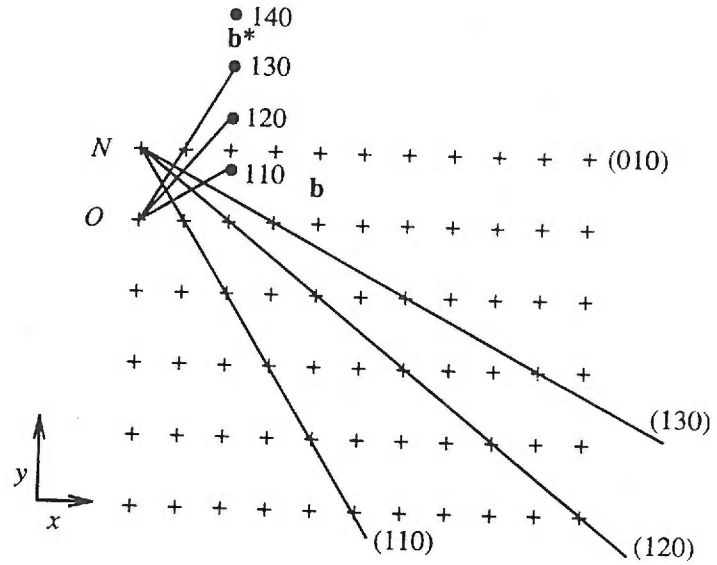
phenomenon is known as constructive interference. For the interplanar distance d, the path difference will be AB+BC=2dsinθ. Then, for constructive interference,

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

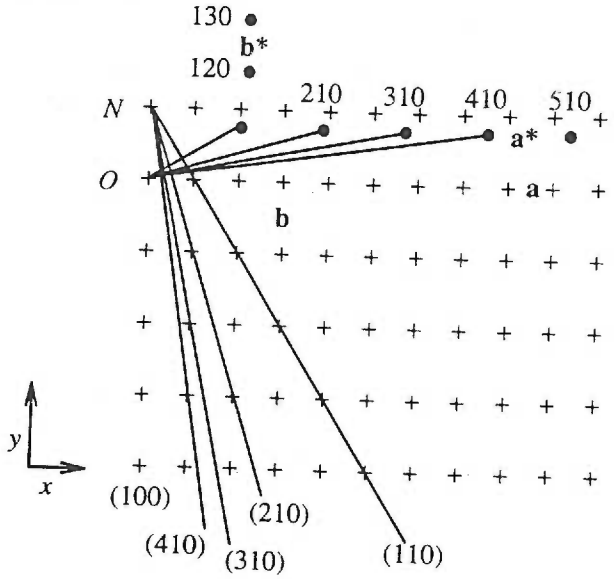
where λ is the wavelength of x-rays and n is an integer. This equation, named Bragg's law, predicts the angle of any diffracted x-ray. The inverse relationship between d and sinθ implies that when crystal planes are close, the diffraction angles will be large whereas when crystal planes are far apart, the angles will be small. By definition, a set of crystal planes hkl produces a reflection hkl on a diffraction pattern. The planes are imaginary, but the atoms are real; thus, the intensity of a given reflection hkl will depend on the distribution of atoms along those hkl planes.

In Bragg's law, sinθ is inversely proportional to interplanar spacing d. But the interpretation of the diffraction pattern would be facilitated if one could substitute the inverse relationship between d and sinθ with a direct one. This is achieved by constructing a reciprocal lattice, based on 1/d. In other words, the dimensions of a reciprocal lattice are inversely proportional to the dimensions of a real lattice.

To construct a reciprocal lattice, imagine an ab section of a two-dimensional lattice (the construction works the same way in three dimensions) with the origin at point O (Figure 2.4, top). Now, through a neighbouring lattice point N, draw one plane from each of the sets (110), (120), (130), and so forth. From the origin, draw a normal to the (110) plane and terminate the normal at a point at the distance 1/d(110), the inverse of the interplanar spacing d110 (Figure 2.4, bottom). This point will be a reciprocal lattice point 110. Now repeat the procedure for planes (120), (130), (140), etc. as well as for planes (210), (310), (410), etc. This operation will create a 2-dimensional reciprocal lattice, but this lattice can be analogously continued in the third dimension. The planes hk0, h0l, and



(*a*)



(*b*)

Figure 2.4. Construction of reciprocal lattice (adapted from Rhodes, 1993). Real lattice points are plus signs (+), and reciprocal lattice points are dots.

0kl correspond to xy, xz, and yz planes; they intersect at the origin O and are called the zero-level planes. Other planes of reciprocal-lattice points are called upper-level planes.

The reciprocal lattice, just like a real crystal lattice, has a unit cell with edges a*, b*, and c*; these reciprocal cell edges are inversely proportional to real cell edges. In orthogonal space groups, where the real unit-cell angles $\alpha = \beta = \gamma = 90^\circ$, the reciprocal unit cell axes a*, b*, and c* lie along real unit cell axes a, b, and c, respectively and are reciprocals of the lengths of corresponding real cell edges a, b, and c. For example, a* would lie along a and a*=1/a. The situation is more complex for nonorthogonal cells because angles different from 90° have to be taken into account.

A reciprocal lattice is a mathematical construction, yet has a physical equivalent in a diffraction pattern. In other words, the diffraction pattern of a crystal lattice is a picture of its reciprocal lattice. Crystallographers use a reciprocal lattice because of its convenience in interpreting the diffraction pattern and relating it to the symmetry of the unit cell.

Ewald construction [Ewald, 1921] illustrates Bragg's law in reciprocal space. A crystal is oriented so that the X-ray beam is parallel to the a*c* plane and line XO is parallel to the beam and passes through the origin O of the reciprocal lattice (Figure 2.5, top). The circle with radius $1/\lambda$ is such that its center C lies on the line XO and the origin O of the reciprocal lattice falls on its circumference. A reciprocal lattice point P lies on this circle. By definition, the angle OPB=90°,

$$\sin(\angle OBP) = \sin\theta = OP/OB = OP/2/\lambda.$$

But, since P is the reciprocal lattice point, $OP = 1/d_{hkl}$, and thus,

$\sin\theta = 1\lambda/2d_{hkl}$, and $2d\sin\theta = 1\lambda$, which is Bragg's law for n=1.

Thus, to bring the reflection hkl into the diffraction position, the crystal must be rotated in such a way that this particular reciprocal lattice point falls on the constructed circle. Ewald construction is not limited to the a*c* plane of the reciprocal lattice and works for a sphere

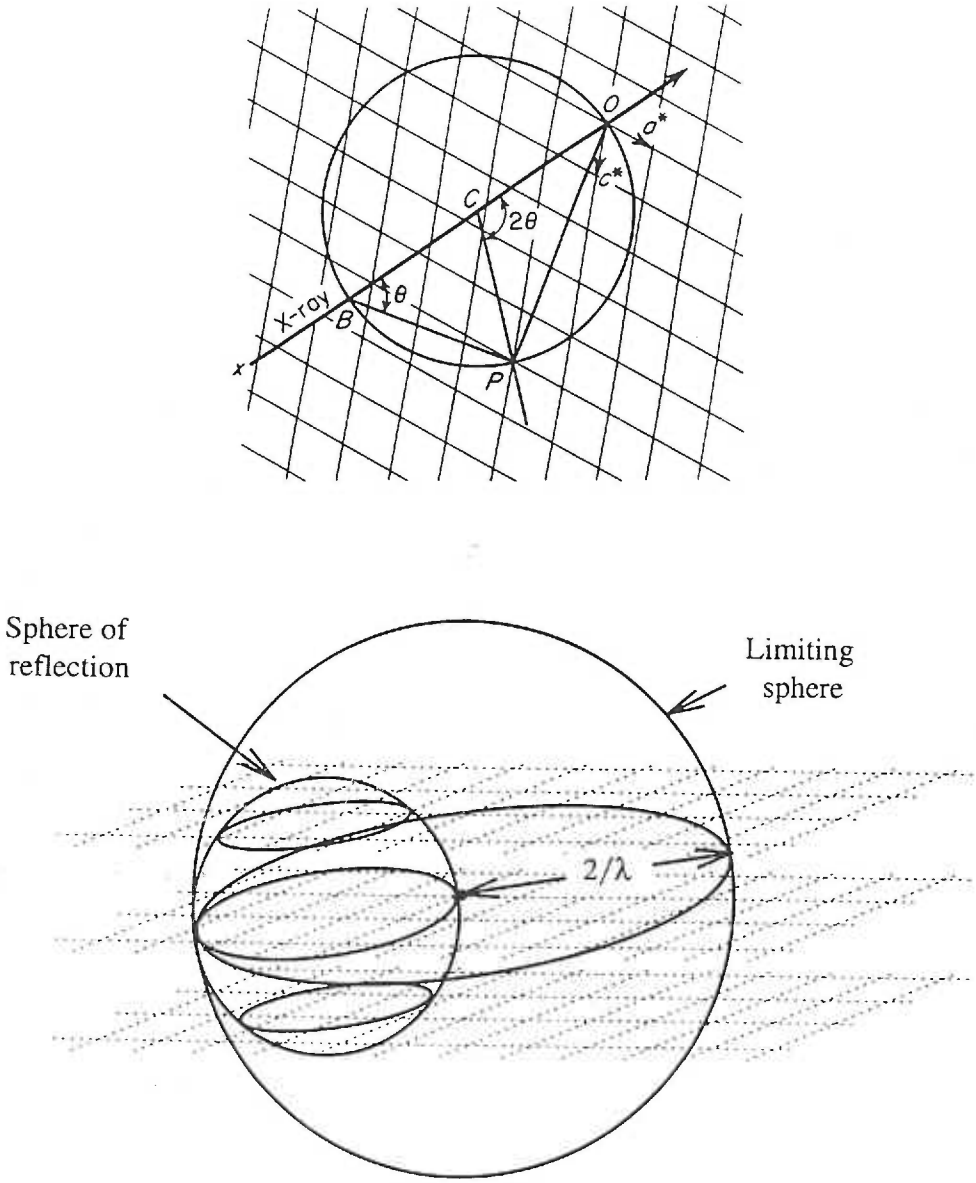


Figure 2.5. Ewald construct describes diffraction in terms of the reciprocal lattice. Top figure: the reciprocal lattice and the sphere of reflection (adapted from Stout and Jensen, 1989). Bottom figure: the limiting sphere (adapted from Rhodes, 1993).

produced by rotating the circle about its diameter OB. The generated sphere is called a sphere of reflection or Ewald sphere. The conditions for Bragg's law are satisfied for any point on the sphere, hence, rotating the reciprocal lattice about its origin will place various reciprocal lattice points on the sphere and the reflections will occur. The reciprocal lattice points that pass through the Ewald sphere are reflections recorded on the detector.

Any reciprocal lattice point within a distance $2/\lambda$ from the origin can be rotated to contact the sphere of reflection. The distance $2/\lambda$ defines the limiting sphere (Figure 2.5, bottom). The number of reciprocal lattice points within the limiting sphere equals the total number of possible reflections that can be produced by rotating a crystal through all possible orientations in the x-ray beam. This number of possible reflections N can be calculated using the equation (for primitive lattices):

$$N = \frac{4/3\pi(2/\lambda)^3}{V_{\substack{\text{reciprocal}\\ \text{cell}}}} = 33.5 \frac{V_{\text{direct cell}}}{\lambda^3}.$$

N depends on the size of the unit cell and the wavelength of the incident radiation λ.

The Bragg's law and the concepts of reciprocal lattice and Ewald sphere are useful to determine the geometry of the data collection. However, they do not reveal the contents of the unit cell that, ultimately, a crystallographer intends to determine.

Structure factor equation.

X-rays are electromagnetic waves, therefore, every diffracted x-ray and the reflection it produces can be described by a Fourier series called a structure-factor equation. The computed sum of the series for a reflection hkl is called a structure factor F_{hkl}. The structure factor is a wave created when many individual waves f_A, f_B, f_C, etc. diffracted by individual atoms A, B, C, etc. are added. In other words, every atom in the structure contributes to every reflection in the diffraction pattern and vice versa. This can be seen from the equation:

$$F_{hkl} = f_A + f_B + f_C + \ldots = \sum_{j=1}^{N} f_j \exp[2\pi i(hx_j+ky_j+lz_j)],$$

where hx_j, ky_j, and lz_j are the fractional coordinates of the jth atom. However, in a crystal, the actual diffractors of x-rays are electron clouds in molecules. Therefore, previous equation can be rewritten to show that each volume element of electron density in the crystal unit cell contributes to the structure factor F_{hkl}:

$$F_{hkl} = \iiint_{xyz} \rho(xyz) \exp[2\pi i(hx+ky+lz)]\, dxdydz = \int_{V} \rho(xyz) \exp[2\pi i(hx+ky+lz)]\, dV.$$

Thus, we can see that the structure factor F_{hkl} is a Fourier transform of the electron density. The Fourier transform is very important in crystallography because it describes precisely the mathematical relationship between an object (a crystalline array of molecules and its electron density) and its diffraction pattern (reflections).

Because the Fourier transform is a reversible operation, an electron density can be represented as the Fourier transform of the structure factors:

$$\rho(xyz) = 1/V \sum_{h}\sum_{k}\sum_{l} F_{hkl} \exp[-2\pi i(hx+ky+lz)].$$

This last equation tells us how to convert the crystallographic data into the electron density $\rho(xyz)$ and to obtain the image of the unit cell contents. Mathematically, we need to measure structure factors, construct their Fourier series and calculate the value of the electron density function ρ at every point xyz.

2. Solving a crystal structure: a practical approach.

Determination of the crystal structure consists of several following steps:

1. Growth of diffraction-quality crystals and their preliminary characterization.
2. Measurement and processing of the intensity data.

3. Calculation of phases.
4. Interpretation of the electron density maps and refinement of a molecular model.
5. Validation of the model.

Crystallization.

The first requirement for the structure determination is to grow suitable crystals: no crystals – no protein structure. The idea behind crystallization trials is to force protein molecules out of their favourable state in solution and into an ordered crystalline state. Solubility of protein molecules in a solvent depends on interactions of the molecule with water, and consequently, low solubility is achieved by dehydrating protein molecules and by modifying properties of the solvent with precipitants. Precipitants usually include salts, polyethylene glycols of various lengths, and organic solvents, such as 2-methyl-2,4-pentane-diol (MPD) and isopropanol [Blundell and Johnson, 1976; McPherson, 1985a, and 1985b]. Different precipitants work by different mechanisms. Polyethylene glycols perturb the structure of water excluding protein from the water phase whereas organic solvents lower the dielectric constant of the medium. Salts have two effects: they dehydrate protein molecules by competing for water and they increase the electrostatic shielding.

All these factors create a supersaturated solution of a protein macromolecule and trigger nucleation - spontaneous formation of nucleation centers – followed by crystal growth. However, if the saturation is reached too quickly or with an extremely high concentration of precipitants, the protein molecules might aggregate to an unwanted amorphous precipitate. So, the strategy used in crystallization is to guide the protein carefully toward minimum solubility. Additional variables affecting crystallization are pH, temperature [McPherson, 1985c], and concentration of the macromolecule. Finally, the protein should be pure since even minor contaminants can interfere with crystallization - hinder nucleation or growth or increase the irregularity of crystals.

We know very little about how macromolecules behave in solution, and thus, crystallization is, perhaps, the most challenging step in crystallographic analysis, which proceeds through trial-and-error. Routinely, several hundred initial trials have to be set up to obtain protein crystals. Commonly used methods include a sparse matrix approach [Jandarik and Kim, 1991] or incomplete factorial approach [Abergel et al., 1991]. These methods allow one to quickly screen the effects of a wide range of pH (4.6 to 10.5), ionic strength, counter ions, temperature (4° C and 21° C), and a multitude of precipitating agents.

In a typical macromolecular crystallization experiment, a small aliquot of the precipitating solution is added to an equal aliquot of a protein solution on a siliconized cover slip. The mixture is inverted and sealed over a reservoir containing 1 mL of the precipitating solution. Over time, water diffuses from the protein droplet to the reservoir, which is approximately twice the ionic strength of the drop precipitant; this process slowly increases the concentration of precipitant in the drop to the critical point when crystals form. The results of crystallization experiments are monitored with a microscope and crystals may appear after a period ranging from several hours to several months.

Once small crystals are obtained, the crystallization conditions are refined until diffraction-quality crystals grow reproducibly. During this fine-tuning process, sometimes a small molecule effector, such as a transition metal ion [Trakhanov and Quiocho, 1995] or non-ionic detergent, is included to optimize the crystallization process. For example, in this work, Mn^{2+} ions were crucial for the production of data-quality crystals of apo BRC protein while a small amount of MPD was required for the crystallization of BRC-TPP complex.

Ultimately, the diffraction quality is the most important characteristic of a protein crystal. While it is possible to determine a fairly good model of a protein to 3 Å resolution, higher resolution data provide a more accurate and detailed structure of a given macromolecule. Often, diffraction depends on the crystal volume, that is, the bigger the

crystal, the better it diffracts X-rays. For a crystallographer, it is thrilling to find the first shower of microcrystals of a new protein but very frustrating not to get crystals to grow to a substantial size. Often, large crystals are obtained simply by increasing the drop volume. However, when this approach fails, a crystallographer can retreat to other techniques to enlarge the crystal volume, e.g., seeding fresh protein drops with microcrystals or setting-up batch crystallization experiments.

Preliminary characterization of crystals.

Once crystals have been grown, the goals of the preliminary characterization are to find how well the crystal diffracts, to determine the symmetry and parameters of the unit cell, and to deduce the space group.

The resolution limit of a crystal as well as subsequent data collection is commonly done using electronic detectors. The most widespread source of x-rays is a rotating anode. X-rays are produced when electrons accelerated by high voltage bombard the metal anode. The commonly used radiation, λ=1.5418 Å, is emitted by copper anode.

First, a crystal has to be mounted in a sealed glass capillary, fixed on the goniometer and translationally aligned. Next, the goniometer carrying crystal is mounted between the x-ray source and the detector. Then its diffraction pattern is recorded and evaluated. If a crystal diffracts beyond 3.0 Å, it is worth it to proceed with the structure determination. Otherwise, one must either improve the diffraction quality of the existing crystals or obtain a different crystal form.

Modern detectors also provide automatic tools to determine the space group, but the most reliable way to determine the internal symmetry and to deduce the space group of a crystal is to examine its reciprocal lattice planes. Precession photography is the only method that gives an accurate and reliable information about the reciprocal lattice, because it gives the undistorted picture of the reciprocal lattice planes.

Determination of the unit cell dimensions and the space group by precession photography.

In the precession camera, the detector is a film and the camera is constructed so that it produces an undistorted image of reciprocal lattice planes. To obtain this undistorted picture, the film must stay parallel to the plane. In Fig 2.6, top, this is a plane through the origin, a zero-level plane. If the crystal and plane do not move, this plane is recorded on film as a circle corresponding to the intersection of the lattice plane and the Ewald sphere. However, if the normal to the film and the normal to the reciprocal lattice plane are rotated around the direction of the x-ray beam, this motion causes the circle of intersection to rotate in the reciprocal lattice plane. (The reciprocal lattice rotates when one rotates the crystal). As a result, a circular part of the lattice is recorded on film. But if no precautions are taken, reflections from other reciprocal lattice planes will be recorded. Placing an annular screen between the crystal and the film eliminates these reflections (Figure 2.6, bottom). A more detailed description of alignment and setting up the precession can be found in the literature [Blundell and Johnson, 1976; Drenth, 1994].

Finally, the precession photograph shows an undistorted plane of any level, e.g., 0kl, 1kl, or h0l, also called a zone. The zero-level zones can be used to measure the cell edges. For an orthogonal space group, a cell edge a can be calculated by a formula: $a = \lambda * \Delta / F$, where F is crystal to film distance, λ is the x-ray wavelength, and Δ is the distance from the film origin to the reflection 100. To measure all three cell edges, one needs to get the precession photograph of at least two zones.

Finally, one determines the crystal symmetry by examining the symmetry of the reciprocal lattice. Space groups are determined by examining the diffraction pattern of a precession photograph of a zone and noting any symmetry and any oddities. Certain symmetry elements in the unit cell, such as screws or lattice centering, cause specific reflections to be absent from the diffraction pattern. This phenomenon is known as systematic absences. For example, in the tetragonal space group, there are three possible types of screw axis: 4_1 and 4_3 with only every fourth reflection present and 4_2 with only

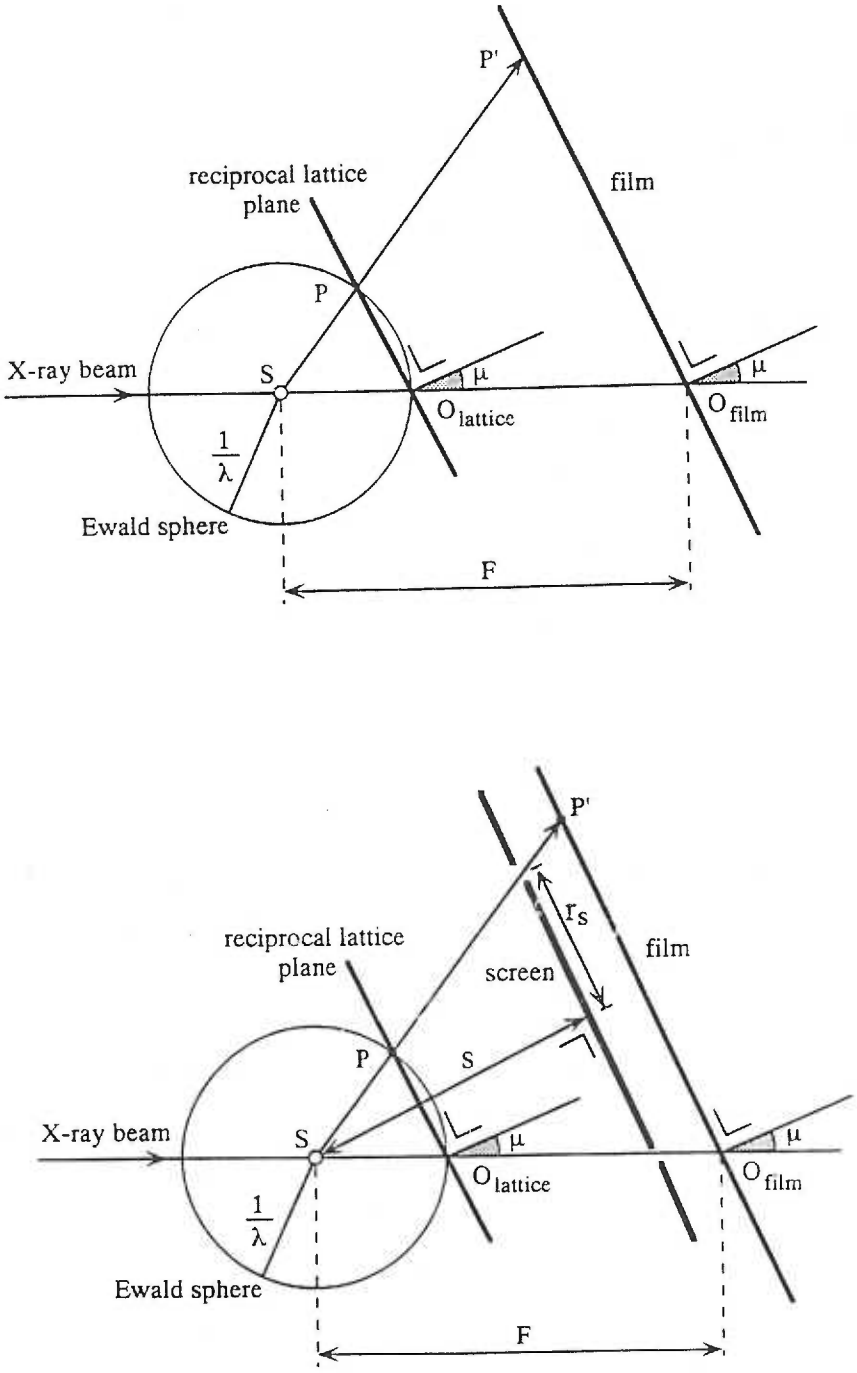


Figure 2.6. Precession photography. a) A precession camera is recording a zero plane of the reciprocal lattice; μ is the precession angle; the crystal to film distance is F. b).The annular screen allows only the zero layer to be recorded on the film. Figures are adapted from Drenth, 1994.

every second reflection present. As described in this work, apo BRC protein crystallizes in a tetragonal space group. A precession photograph of the 0kl zone shows only every fourth spot present along the c-axis, indicating that apo BRC crystals have a 4_1 or 4_3 screw along c-axis but not a 4_2 screw. (The ambiguity was resolved at later stages of analysis in favour of 4_1.) As another example, if the crystal lattice is centered, this will cause the appearance of a 'diamond-shaped' diffraction pattern. More detailed information about the systematic absences for each space group can be found in the International Tables for Crystallography [Hahn, 1983].

Data collection.

Once the space group and unit cell parameters a, b, c, α, β, and γ have been determined, one may confidently collect the intensity data. The essence of the data collection strategy is to record the intensity I_{hkl} of every unique reflection hkl at least once. However, to decrease random error, every reflection should be collected several times. The strategy of the data collection is guided by the knowledge of crystal symmetry. Looking at the Ewald sphere, one may think that the data collection has to cover all 360°, but when the crystal displays high symmetry, certain crystal orientations will allow to measure the same reflections.

Knowing the space group of the crystal, a crystallographer can deduce the space group of the reciprocal lattice and thus know which orientations will produce identical data. Any reciprocal lattice has a symmetry element called inversion centre at the origin. This feature causes the intensity of any hkl reflection to be equal to the intensity of -h-k-l reflection (Friedel's law). This automatically means that half of the reflections in the reciprocal cell are redundant and the data collection has to cover 180° instead of 360°.

Other symmetry elements cause further reduction in the total angle of the data collection. The presence of the inversion centre reduces 230 space groups to 11 Laue groups for the reciprocal space. For each Laue group, one computes the fraction of

reflections that are unique. As has been mentioned earlier, knowledge of the crystal symmetry greatly reduces the number of reflections that must be collected.

The result of the data collection is a list of intensities with assigned indices hkl corresponding to the position in the reciprocal lattice. For every reflection hkl, the intensity I_{hkl} has to be converted into a structure factor amplitude $|F_{hkl}|$. From these structure factor amplitudes, crystallographers extract structural information.

Phase determination methods.

The electron density in a crystal can be obtained according to equation:

$$\rho(xyz) = 1/V \sum F_{hkl} \exp [-2\pi i(hx+ky+lz)].$$

The structure factor F_{hkl} for a given reflection from a crystal can be represented by its amplitude $|F_{hkl}|$ and phase α_{hkl}: $F_{hkl} = |F_{hkl}|\exp(i\alpha_{hkl})$.

However, from the diffraction pattern, one can measure only the intensity I_{hkl} of the structure factor. Because I_{hkl} is proportional to $|F_{hkl}|^2$, one can calculate the amplitude $|F_{hkl}|$ of the structure factor. But the phase α_{hkl} cannot be recorded, hence one cannot determine the structure factor F_{hkl}. This is known as "the phase problem" of crystallography.

Several methods are routinely used to solve the phase problem. A crystallographer's choice depends on how much structural information is available about a given protein. The multiple isomorphous replacement method (MIR) does not need any *a priori* structural information about a protein and allows to solve a protein structure *de novo*; but heavy atoms must be attached to protein molecules in the crystal. In contrast, the molecular replacement method (MR) requires the structure of a homologous molecule but no derivatives. Next, the molecular modification method (also known as the difference Fourier method) is routinely used to solve structures of a modified protein–a mutant or a ligand complex–that crystallizes isomorphously with the wild-type protein. Finally, the multiple-wavelength anomalous-diffraction method (MAD) is efficient for determining *de*

novo structures but requires strong anomalously-scattering atoms in the protein. The first three methods have been used in the present work and will be discussed in more detail.

Multiple isomorphous replacement.

MIR method is the most frequently used to determine *de novo* crystal structures. This method entails the incorporation of "heavy" atoms –atoms with high atomic numbers - into a protein and the subsequent determination of the coordinates of those heavy atoms in the crystal unit cell. In this work, the method was used to solve the structure of apo BRC protein.

Preparation of heavy atom derivatives.

Most frequently, crystals of protein are soaked in solutions containing heavy metal ions or their complexes. Commonly used heavy metals are Hg, Au, and Pt, which are capable of binding to specific sites in the protein. For example, Hg^{2+} ions react with cysteines while complexes like $AuCl_4^-$ and $PtCl_4^{2-}$ preferentially target methionines and histidines. The conditions that give specific binding are usually found by trying different soaking times at different concentrations of the heavy metal ions at different pHs. In addition, heavy atoms can be introduced into a protein by incorporating amino acids during biosynthesis. Often, selenomethionine or telluromethionine are used to replace methionine.

Derivative crystals have to satisfy three criteria to be useful in the phase determination. First, they have to be isomorphous with the native crystals. That means that heavy atom should not disturb the packing or the conformation of macromolecules in a crystal. The unit cell dimensions are sensitive to such changes and can signal nonisomorphism of a crystal. Second, a heavy atom must produce observable changes in reflection intensities. These changes are key to phase determination and must be reasonably large to be measured accurately. Finally, the derivative crystals must diffract

reasonably well. The quality of a potential heavy-atom derivative can be quickly checked by calculating an R_{iso} value:

$$R_{iso} = \frac{\Sigma \, ||F_{PH}| - |F_P||}{\Sigma \, |F_P|} \times 100\%.$$

A good derivative should have an R_{iso} between 14 and 25%. An R_{iso} above 30% indicates nonisomorphism whereas an R_{iso} below 10% means that the crystal has not been derivatized.

Locating heavy atoms using Patterson and Fourier methods.

The most powerful tool in determining the heavy atom coordinates is the Patterson function P(uvw):

$$P(uvw) = \Sigma \, |F_{hkl}|^2 \exp[-2\pi i(hu+kv+lw)].$$

This function is the Fourier transform of $|F_{hkl}|^2$ and is similar to the electron density function $\rho(xyz)$ – a Fourier transform of F_{hkl}. The Patterson function does not include phases.

The coordinates uvw on the Patterson map are analogous to the coordinates xyz on the electron density map. However, in an electron density map, peaks correspond to locations of atoms. In the Patterson map, peaks show locations of interatomic vectors (Figure 2.7). Patterson maps are very noisy and hard to interpret – there are more vectors between the atoms than atoms themselves. For N atoms there will be N×(N-1) Patterson peaks plus a huge peak at the origin due to self-vectors. However, for just a few heavy atoms in the unit cell, Patterson map can be interpreted.

The basic idea is to separate the heavy atom contribution from the protein. For every reflection in the heavy atom derivative, the structure factor is a sum of the structure factor for the native protein and the heavy atom: $F_{PH} = F_P + F_H$.

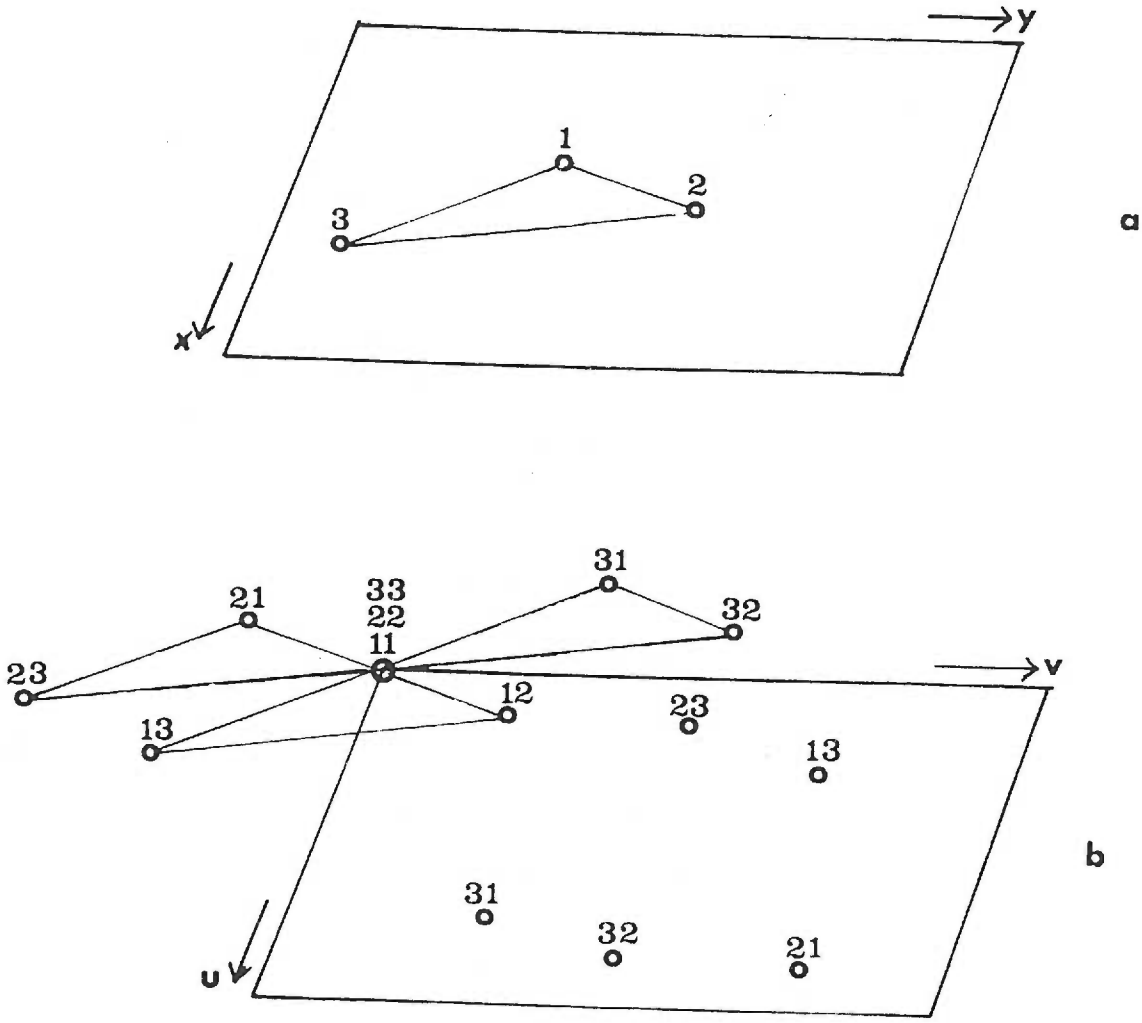


Figure 2.7. A simple structure and its Patterson function (adapted from Blundell and Johnson, 1976).

Thus, the contribution of heavy atom is defined as $F_H = F_{PH} - F_P$. However, we know only the amplitudes $|F_{PH}|$ and $|F_P|$, not the complete structure factors F_{PH} and F_P. To obtain the Patterson function for the heavy atom in the derivative crystals, one has to construct a difference Patterson function, using as $(\Delta F_{hkl})^2$ as Fourier coefficients:

$$P(uvw) = \sum |\Delta F_{hkl}|^2 \exp[-2\pi i(hu+kv+lw)],$$

where $(\Delta F_{hkl})^2 = (|F_{PH}| - |F_P|)^2$. Each (ΔF_{hkl}) is derived from the same hkl reflection in the native and derivative data sets and represents the contribution of a heavy atom to the structure factor F_{PH}. This will be true only in the centric zones, where F_{PH} and F_P are collinear and $|F_H| = ||F_{PH}| +/- |F_P||$. In general, this will not be true for acentric reflections, where F_{PH}, F_P, and F_H are not collinear, except for most of the largest $||F_{PH}| - |F_P||$ differences. So, we can use the ΔF for centric and the largest ΔF for acentric reflections to locate heavy atoms.

Unit cell symmetry simplifies the search for peaks (yet another advantage of knowing the symmetry of a crystal). Due to the symmetry, vectors connecting the symmetry related atoms in the unit cell fall on special planes called Harker sections. By correlating vector peaks from different Harker sections, one can determine unambiguously two or all three coordinates of heavy atoms.

Often, due to the noise on a Patterson map, only the high-occupancy heavy-atom sites can be found. In this case, one can locate minor heavy-atom sites using difference Fourier techniques.

$$\Delta\rho(xyz) = 1/V \sum_{h\,k\,l}\sum\sum |F_{PH}| - |F_P| \exp[-2\pi i(hx+ky+lz) + i\alpha_P(hkl)],$$

where the initial estimate of protein phases α_P is calculated from coordinates of strong heavy-atom sites, as described below. Moreover, once protein phases α_P are estimated from one derivative, difference Fourier method can be used to locate heavy-atom sites in

other derivatives. This method is known as cross-phase difference Fourier. The phases α_P are calculated from the first derivative and $|F_{PH}|$ are used from another derivative. Conveniently, heavy atom solutions from all the derivatives have the same origin. In all cases, heavy-atom sites of a derivative identified through difference Fourier techniques should be checked against the Patterson map of that derivative. This is important since the phases α_P calculated from the heavy-atom peaks in the first derivative, often reproduce those peaks as the ghost peaks on difference Fourier maps.

Having solved one or more derivative data sets for heavy atom positions, the next step is to refine the positions and to evaluate the quality of a derivative. Heavy atom parameters can be improved by one of two methods. In the first, the heavy atom parameters are refined by minimizing the function $[F_{H(obs)} - F_{H(calc)}]^2$, where F_H for each reflection hkl is calculated from the coordinates of heavy atoms in the unit cell. The strength of this scheme is its independence of other derivatives. However, if protein phases are estimated from a heavy atom derivative, they can be used in the refinement of a new derivative. The used minimized function will be $(F_{PH(obs)} - F_{PH(calc)})/E$, where E is the standard error of the derivative structure factors.

Obtaining protein phases from heavy atom data.

The structure factors F_{PH}, F_P, and F_H are related by the following equation:

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

Ultimately, one is interested in F_P but knows only its amplitude $|F_P|$ and not the phase α_P. $|F_{PH}|$ and $|F_P|$ can be measured experimentally, and $|F_H|$ and α_H can be computed from heavy atom coordinates. How then can the unknown protein phase α_P be obtained from the experimental data? Harker proposed the geometric analysis [1956] of this problem.

For convenience, we can start with a rearranged form of the above equation

$F_P = F_{PH} - F_H$. The structure factor vectors are then constructed using the two-dimensional Argand diagram (Figure 2.8) where a structure factor F is drawn as a vector of the length |F| from the origin in such way that the angle between the vector F and the horizontal axis is equal to the phase α. The Harker construction is accomplished in three steps (Figure 2.9, top). First, we can measure $|F_P|$ but not α_P, so the tip of the F_P vector must lie somewhere on the circle with radius $|F_P|$ and centered on the origin. Second, we know both $|F_H|$ and α_H, we can draw a vector $-F_H$. Finally, we can measure $|F_{PH}|$ but not α_{PH}, thus, the tip of the F_{PH} vector must lie somewhere on the circle with radius $|F_{PH}|$. However, in this case, the circle is centered on the tip of the $-F_H$ vector, because according to the equation above, the vector F_{PH} has to be added to the vector $-F_H$.

From the resulting diagram, 2 circles intersect at two points. Therefore, the protein phase α_P will have 2 possible solutions symmetrically displaced around the heavy atom phase α_H (Figure 2.9, top). The same solution in the algebraic form is:

$$\alpha_P = \alpha_H +/- \arccos \frac{|F_{PH}|^2 - |F_P|^2 - |F_H|^2}{2\,|F_P|\,|F_H|}.$$

But which solution is correct? This dilemma is referred to as the phase ambiguity problem. To resolve the phase ambiguity, a second different heavy atom derivatives is necessary. Using data from the second isomorphous derivative, one can generate the Harker diagram for the second derivative (Figure 2.9, bottom). The second derivative will also indicate 2 possible solutions, symmetrically disposed about its heavy atom vector. Only now, if the two diagrams are superimposed, only the correct solution will be at the intersection of three circles. Clearly, in two derivatives, heavy atoms must bind to different sites on a protein; otherwise, vectors F_{H1} and F_{H2} will be collinear.

The Harker method assumes that the structure factors are perfectly measured and that the derivative crystals are isomorphous with the native crystals. In practice, however,

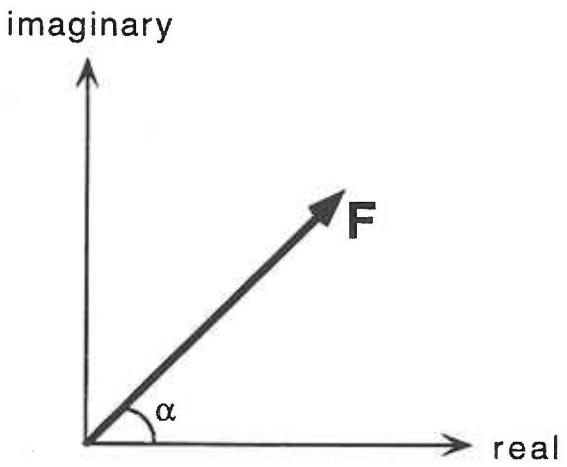


Figure 2.8. Argand diagram.

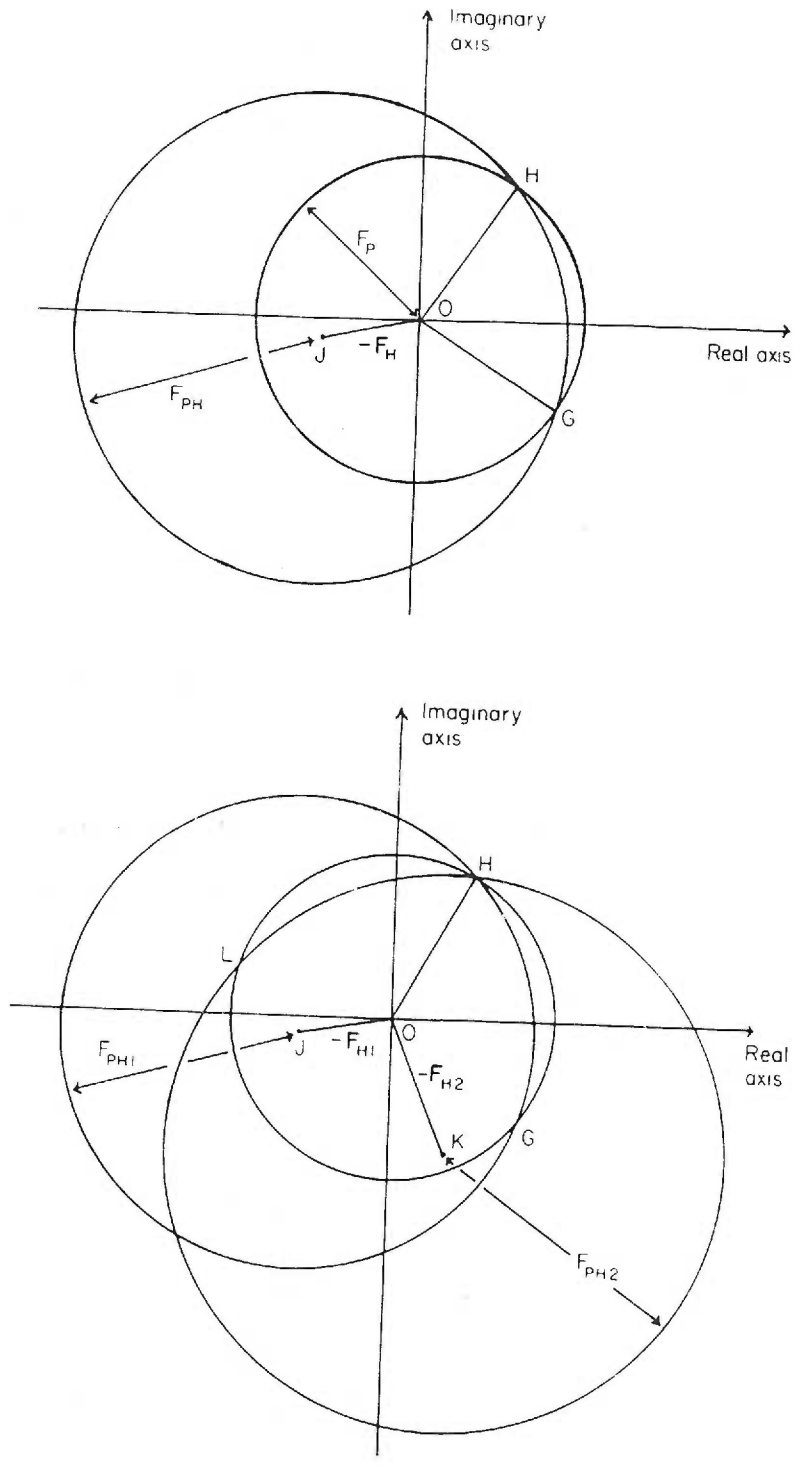


Figure 2.9. The Harker construction for phase calculation by the method of single (top) and double (bottom) isomorphous replacement. Figures are adapted from Blundell and Johnson, 1976.

the circles in the Harker diagram might not intersect exactly or often might not intersect at all. This could be caused by experimental error in the measurements, incorrect heavy atom positions or by non-isomorphism. Therefore, to know how reliable phases are, one should incorporate the error model for the breakdown of the equation $F_{PH} = F_P + F_H$.

Treatment of errors.

Blow and Crick [1959] have proposed a model in which all the error lies in the value of $|F_{PH}|$ with a Gaussian distribution. They introduced the "lack of closure" error $\varepsilon(\alpha_P)$, which shows how well the equation is satisfied for any given phase angle:

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

where $|F_{PH}|_{calc}$ is obtained from the measured $|F_P|$, computed F_H, and assumed α_P. The term "lack of closure" can be illustrated in the absence of the error, the three vectors F_{PH}, F_P, and F_H would "close" to form a triangle (Figure 2.10, top). Any error would prevent such "closure" (Figure 2.10, bottom). For every reflection hkl, the probability P for a given phase α_P is then related to the lack of closure by

$$P(\alpha_P) = \exp\left[\frac{-\varepsilon(\alpha_P)^2}{2E^2}\right],$$

where E is the "standard error" - the breadth of Gaussian distribution for the error ε. E can be estimated from the $\varepsilon(\alpha_P)$ values for centric reflections, because α_P for the centric reflections is either 0 or π. Basically, one calculates the value of $P(\alpha_P)$ in the range 0 - 2π to determine the most probable protein phase, i.e. the one that minimizes the lack of closure.

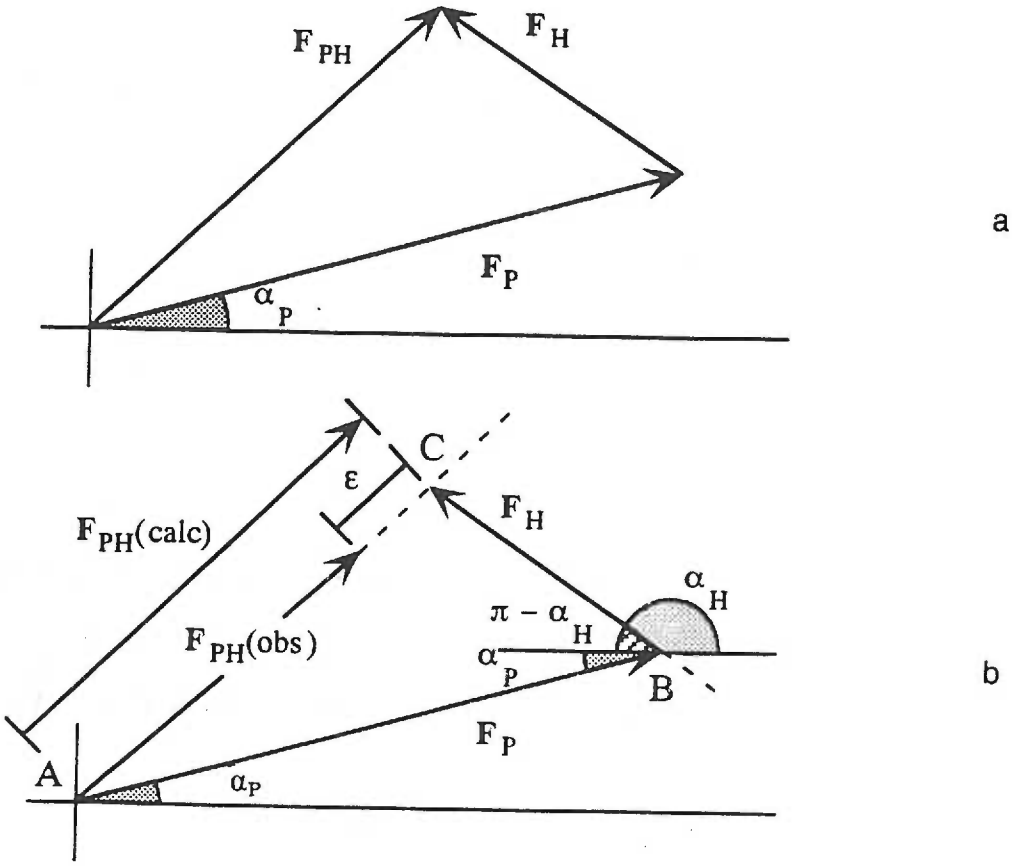


Figure 2.10. a) The ideal isomorphous situation in which the vector triangle $F_{PH} = F_P + F_H$ closes exactly. Normally, this is not true and the observed and calculated values of $|F_{PH}|$ differ by the lack of closure error ε, as shown in b). Figure is adapted from Drenth, 1994.

An alternative model has been proposed by Hendrickson and Lattman [1970], who assumed that all the error lies in the derivative intensity, proportional to $|F_{PH}|^2$, also with a Gaussian distribution. The corresponding "lack of closure" error $\delta(\alpha_P)$ is defined as:

$$\delta(\alpha_P) = (|F_{PH}|_{obs}^2 - (|F_{PH}|_{calc}^2),$$

and the probability is

$$P(\alpha_P) = \exp\left[\frac{-\delta(\alpha_P)^2}{2D^2}\right],$$

This error model is convenient because the probability distribution can be encoded in terms of only four constants:

$$P(\alpha_P) = K\exp[A\cos\alpha_P + B\sin\alpha_P + C\cos(2\alpha_P) + D\sin(2\alpha_P)],$$

where K is the scaling factor. The constants A, B, C, and D can be calculated not only from the heavy atom data but also from other sources such as a partial structure of the molecule. To combine probabilities from different sources (for instance, different heavy atom derivatives), one can simply add the corresponding coefficients A, B, C, and D. This means that every time a new derivative is obtained, one does not have to recalculate the entire probability distribution.

The most probable phase is the correct estimate of α_P only for the monomodal distribution of $P(\alpha_P)$ (Figure 2.11). When the distribution is bimodal (Figure 2.11), the most probable phase is biased towards the uncertain phase. The centroid "best" probability distribution is better, because the electron density maps computed with $\alpha_P = \alpha_{best}$ have the least mean square error over the entire unit cell. The best value of F_{hkl} is

$F_{hkl}(best) = m|F_{hkl}| \exp(i\alpha_{best})$, where m is defined as

$$m = \frac{\int P_{hkl}(\alpha) \exp(i\alpha) d\alpha}{\int P(\alpha) d\alpha}$$

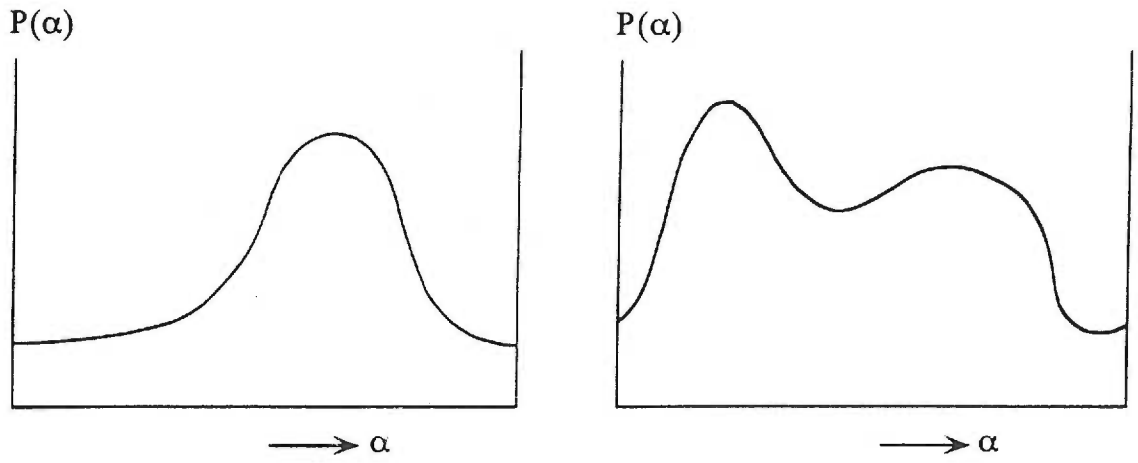


Figure 2.11. Two examples of the total probability P(α) for the phase angle α of a reflection as derived from more than one derivative.

and can assume any value between 0 and 1. If the probability is sharp, m will be close to 1, but if the probability is uniform from 0 to 2π, m will be close to 0 and indicates a flat – useless – distribution of phases, where all phases are equally probable. m is called the "figure of merit" and is equivalent to a weighting function. The electron density calculated with m|F| and α_{best}, is expected to have a minimum mean square error from the true electron density.

For phase calculation, the probability function $P(\alpha_i)$ is computed in fixed intervals, e.g., 5 degrees, at angles α_i. The figure of merit m is calculated from the expression

$$m\cos(\alpha_{best}) = \frac{\Sigma P(\alpha_i)\cos\alpha_i}{\Sigma P(\alpha_i)} .$$

The error in the phase angle at a given angle α_i is defined as $\Delta\alpha_i = (\alpha_{best} - \alpha_i)$. If the origin is taken at the center of gravity, $\alpha_{best} = 0$ and $|\alpha_i| = |\Delta\alpha_i|$. The, $m = \langle\cos\Delta\alpha_i\rangle$. For m=1, the error in the phase angle is 0, whereas for m=0.74, the error is 42°. Thus, m serves as a valuable indicator of the phase quality.

In practice, in addition to figure of merit, other statistical values can be used to judge the quality of a derivative and of phases. One of the most important numbers in phasing statistics is the phasing power or $\langle F_h\rangle/\langle E\rangle$, the size of heavy atom amplitudes over the error. The phasing power is calculated by the formula:

$$\left[\frac{\Sigma |F_H|^2}{\Sigma |E|^2}\right]^{1/2}, \text{ with } \Sigma|E|^2 = \Sigma\{|F_{PH(obs)}|-|F_{PH(calc)}|\}^2$$

and simply indicates a "signal to noise" ratio for a particular derivative. This number should usually be greater than 1 for the derivative to contribute to phasing. To increase the phasing power, one may either increase the heavy atom contribution F_H by using longer

soaking times and higher concentrations of heavy metals in the soaks or decrease the error E by improving the solution and collecting better data. The phasing power decreases with resolution because the scattering of heavy atoms falls off with resolution.

Molecular Replacement.

In the molecular replacement method, the initial estimates of phases for a new protein can be calculated from the phases of a model – a known homologous protein. Homology in the amino acid sequence indicates whether a model is suitable. The initial phases are calculated by placing a model of the known protein in the unit cell of the unsolved protein. Such placement requires knowing the correct orientation and position of the molecule. In this work, molecular replacement was used to solve the structure of BRC-TPP complex using the structure of apo BRC as an initial model.

Patterson methods to solve molecular replacement problems.

Molecular replacement using Patterson methods involves 2 steps: rotation and translation. In the rotation step, one has to find the spatial orientation between the known and the unknown molecules. In the following translation step, one calculates the translation that must superimpose the model molecule and the unknown molecule.

The aforementioned Patterson map is essentially a vector map: vectors between the atoms show up as vectors from the origin to peaks in the Patterson map. If the pairs of atoms belong to the same molecule, the vectors between them will be relatively short (less than the longest dimension of the molecule); the peaks corresponding to these so-called self-vectors show up close to the origin in the Patterson map. This region should be very similar for homologous molecules in different crystal systems, only with a rotation difference. Briefly, one looks for a maximum of the rotation function, R, which is calculated as an overlap of two Patterson functions:

$$R(\alpha\beta\gamma) = \int_{U} P(u) \times P_r(u_r)\, du,$$

where P(u) is the Patterson function for the unknown molecule, $P_r(u_r)$ is the rotated version of the Patterson function for the model molecule, and U is the volume in the Patterson map where self-vectors are located. A detailed description of rotation search can be found in many crystallographic textbooks [Blundell and Johnson, 1976; Drenth, 1994].

Once the orientation of the unknown molecule is resolved, a translation is required to overlap the model with the unknown molecule. One computes a translation function that correlates the set of cross-Patterson vectors for the model structure $P_{calc}(u,m)$ with a set of the observed Patterson function $P_{obs}(u)$:

$$T(m) = \int_{V} P_{obs}(u) \times P_{calc}(u,m)\, du.$$

T(m) should be at a maximum when the model is in the best agreement with the observed data. In an alternative trial-and-error approach, the model molecule is moved through the asymmetric unit and structure factors are calculated and compared with the observed structure factors.

Several molecular replacement programs such as MERLOT [Fitzgerald, 1988] and AmoRe [Navaza, 1994] implement Patterson methods to solve structures. The solutions of the rotation and translation function are not always trivial to find. Often, it can be necessary to modify the original model. For example, the original model must be truncated by deleting doubtful parts, using one monomer of the original oligomeric molecule, or using just one domain of the model molecule.

Using genetic algorithm to solve molecular replacement problems.

Despite the wide use of Patterson-based molecular replacement methods, a new method based on the use of genetic algorithms is gaining popularity. EPMR (Evolutionary Program for Molecular Replacement) is one of the first programs that find crystallographic

molecular replacement solutions using an evolutionary search algorithm [Kissinger et al., 1999].

Unlike traditional molecular replacement searches, which first determine the three rotational and then the three translational parameters, EPMR directly optimizes all six parameters, in other words, EPMR searches six-dimensional space for the correct solution.

Briefly, the genetic algorithm (GA) is based on the principles of evolution and manipulates populations, strings, mutations, fitness, mating, etc. [Harris et al., 1998]. The goal of GA is to ultimately produce offspring that are optimal with respect to a criterion used to define the quality of the individual members in the population. A string (set) of variables is equivalent to a model in a certain orientation where individual variables correspond to 3 rotation and 3 translation parameters. The fitness of a string is a measure of its quality in respect to an optimized property, which, in our case, is the correlation coefficient between the observed structure factors F_{obs} and the structure factors F_{calc} calculated from the starting model. This criterion is used to discriminate between bad and good members of a population. For example, a structure giving rise to a good agreement between the observed and the calculated structure factors will have a high fitness. A population of strings - trial structures - is made to evolve depending on rules governing mating, mutation and natural selection. The evolution is such that the strings of highest fitness have the best chance to survive and pass their characteristics onto subsequent generations to produce offspring of high quality. One of the advantages of the GA method is that the initial population does not have to be close to global minimum. Another advantage of the GA method is that many different regions of parameter space are investigated simultaneously.

The program EPMR has several stages. First, it generates an initial set of random orientations for the search model and calculates a correlation coefficient for each orientation. Next, a fraction of the highest-scoring orientations is retained and a complete set of new trial orientations is generated by randomly altering the rotation angles and

translations of each surviving solution. Once again, correlation coefficients are calculated and the population is regenerated from the top scoring solutions. The program repeats this procedure for several cycles until a certain criterion (e.g., the correlation coefficient exceeding 0.5) is met. At the end of this evolutionary optimization, a rigid-body refinement of the search molecule orientation and position is performed.

Molecular Modification.

In the method of molecular modification, the structure of the modified protein is very similar to the structure of a wild-type protein, i.e., the modified protein is isomorphous with the wild-type protein. Thus, the structure factors are calculated using the experimental amplitudes and the phases of a homologous structure. One calculates difference Fourier maps with coefficients, 2Fo-Fc

$$\rho(xyz) = 1/V \sum_{h\,k\,l}\sum\sum (2|F_{obs}| - |F_{calc}|) \exp[-2\pi i(hx+ky+lz) + i\alpha_{model}]$$

and Fo-Fc:

$$\rho(xyz) = 1/V \sum_{h\,k\,l}\sum\sum |F_{obs}| - |F_{calc}| \exp[-2\pi i(hx+ky+lz) + i\alpha_{model}],$$

and examines the entire map to detect any change in modified protein. In this work, the Molecular Modification method was used to solve the structures of PurR-hypoxanthine bound to modified *purF* operators using the structure of PurR-hypoxanthine-*purF* operator complex (PDB code *1pnr)* [Schumacher et al , 1994].

Phase improvement.

Because the initial calculation of structure factors and phases is inaccurate, the initial electron density map and, therefore, the model are inaccurate. Insufficient quality of the electron density map might impede an unambiguous tracing of the polypeptide chain. To improve the accuracy of the model, a crystallographer attempts to improve the electron density map by density modification methods, which include solvent flattening,

incorporation of partial structure information, and non-crystallographic symmetry-averaging. Crystallographic papers are replete with examples in which applying the density modification methods greatly improved otherwise uninterpretable electron density maps. In particular, in this work, solvent flattening was indispensable in producing interpretable electron density maps of apo BRC.

Briefly, the method of solvent flattening allows a crystallographer to define the regions of the electron density between those occupied by macromolecules and those occupied by solvent molecules. Highly refined structures of protein show that electron density of the solvent is "flat" or featureless due to the dynamic character of solvent molecules. However, experimental electron density maps are quite noisy and "peaks" of density are abundant throughout these initial MIR maps. If we find the molecular boundary of the protein, we can set the electron density in the solvent region to a small constant value – "flatten" it - and decrease the noise of the map.

The three-dimensional grid is superimposed on the unit cell. At each grid point j, the electron density is replaced by a new value that is proportional to the weighted sum of the densities within a sphere of radius R centered on the grid point j.

$$\rho_j' = \sum_i^R w_i \rho_i \text{ , with } w_i = 0 \text{ for } \rho_i < 0 \text{ and } w_i = 1 - (r_{ij}/R) \text{ for } \rho_i > 0.$$

The summation is over the grid point i within the sphere; r_{ij} is the distance between the grid points i and j. Therefore, if the grid point is surrounded primarily by solvent, the new density ρ_j' will be low whereas for the grid point within the protein, the new density will be high. Structure factors and amplitudes are calculated for this new map. The electron density is calculated with observed structure factors and with the new phases either from the solvent flattening procedure or combining them with the MIR phases [Wang, 1985; Furey and Swaminathan, 1997]. Solvent flattening is most efficient when the solvent content of the crystal is high (>50%).

Another way to improve the existing model is to combine the isomorphous replacement information with phase information from the known part of the structure. The most convenient way to combine phase information from various sources is to add individual probability distributions. The probability function in terms of Hendrickson-Lattman coefficients is

$$P_{partial}(\alpha_P) = N \exp[X \cos(\alpha_P - \alpha_{partial})] = N \exp[\overbrace{X\cos\alpha_{partial}}^{A_{partial}} \cos\alpha_P + \overbrace{X\sin\alpha_{partial}}^{B_{partial}} \sin\alpha_P]$$

where N is a normalizing constant and X accounts for errors [Drenth, 1994].

Interpretation of the electron density maps.

The result of the diffraction experiment is the three-dimensional map of the electron density for a given macromolecule. However, a crystallographer must build an atomic model of this macromolecule to fit the electron density. The model must explain particular features of the map and be consistent with available stereochemical information. This part of the structure determination is primarily manual. In other words, a crystallographer inspects the electron density map and builds the model that would explain particular features of the map. This is usually done using an interactive computer graphics program. All model building in this thesis has been done using O [Jones et al., 1991] on a Silicon Graphics Indigo (SGI) workstation.

At this point, the crystallographer takes advantage of all available information: primary structure of the protein and predicted secondary structure; biochemical data, including all modifications, interacting amino acid groups, etc. The first stage of interpretation involves identifying and building-in secondary structure elements: α-helices and β-strands. Next and one of the most challenging steps in the interpretation is matching

the primary sequence to the map. Since it is very difficult to tell all the side chains apart by their densities, a crystallographer looks for a pattern of large and small side chains that is unique to a given protein. Amino acids with bulky side chains, such as tryptophans and tyrosines, can be easily recognized and greatly alleviate the task of the map interpretation, which resembles solving a jigsaw puzzle.

Refinement.

From the MIR or molecular replacement method, a crystallographer obtains an approximate model of the protein structure, which is generally correct in the overall shape and folding of the molecule. However, because of the lower initial resolution, such a model is inaccurate in the atomic details, and the structure factors calculated from this model agree poorly with the experimentally observed structure factors. The purpose of crystallographic refinement is to allow a closer agreement between the calculated and observed structure factors, which is done by optimizing the function:

$$E_{cryst} = \sum_{hkl} w(hkl)\{|F_{obs}(hkl)| - |F_{calc}(hkl)|\}^2.$$

The agreement between the observed and calculated structure factors can be accessed by the value of the R-factor:

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

For a starting MIR model, the R-factor is usually in the range 35% - 45% and can be as high as 50% whereas for a random acentric structure, R-factor would be 59%. Refinement is the process of adjusting the model to find a closer agreement between the calculated and observed structure factors, and it usually lowers the R-factor to below 25%.

However, because of the relatively low resolution, a macromolecular crystal seldom produces enough diffraction data to refine a structure independently. To compensate for

low resolution of x-ray data, stereochemical restraints are included. The stereochemical data are obtained from the small molecular structures including amino acids and short peptides. Their bond lengths and angles are determined with high precision and are assumed to be valid in proteins [Engh and Huber, 1991]. During refinement, the stereochemical parameters are "restrained" - allowed to vary around a standard value. These restraints are the additional observations. The stereochemical information can be incorporated into the refinement in the form of additional terms in the expression for the minimized function E_{total}:

$$E_{total} = w_A \Sigma E_{cryst} + w_B \Sigma E_{bonds} + w_C \Sigma E_{angles} + w_D \Sigma E_{planar\ groups} + \text{etc.}$$

Every stereochemical term E_i is calculated by estimating the difference between the actual value in the structure and the ideal value $E = (Q_m - Q_i)^2$. The further the actual value is from the ideal value, the higher the energy; w - the weight given to any term - should be about $1/\sigma^2$.

Several different refinement methods are available to the crystallographic community. All refinement in this work was done using the least-squares refinement method as implemented in the software package TNT [Tronrud et al., 1987; Tronrud, 1997]. A typical refinement starts with the rigid-body refinement, in which the core of the molecule is moved as a single unit. Often, this is followed by the rigid-body domain refinement where the molecule is broken into rigid domains that are refined separately. Thus, the rigid-body refinement permits one to account for the gross errors in positions of the whole molecule and its domains. Next, every atom in the model is adjusted by changing the coordinates x, y, and z and the thermal parameter B. At this stage of the refinement, solvent molecules are usually added. A potential water molecule should not make bad contacts with the protein and should be within the hydrogen bonding distance (2.7 to 4.0 Å) from the hydrogen bond donors and/or acceptors.

Validation of the model.

During model building and refinement, a number of errors can "creep" into the protein model. To ensure the correctness of work, crystallographers have relied on the R_{factor}, which gives an overall measure of how well the final model fits the experimental diffraction data. However, sometimes, the R_{factor} can be made arbitrarily low if the number of parameters describing the model is too high or not properly restrained. To improve the situation, Brünger introduced a statistical cross-validation parameter R-free, which is unbiased by the refinement process [1992]. For cross-validation purposes, 10% of the data is excluded from the refinement and kept as a reference or "test" set. Only data in the working set are used in the refinement while the test data are used to calculate R-free. The R-free value measures the degree to which the model predicts the diffraction data for the set that has been excluded from the refinement process.

$$\text{R-free} = \frac{\sum_{hkl \subset T} ||F_{obs}| - |F_{calc}||}{\sum_{hkl \subset T} |F_{obs}|},$$

where hkl⊂T means that all reflection belong to the test (T) set. R-free is a global statistic very sensitive to gross errors in a model, and it has been shown to correlate the accuracy of atomic models. The models with serious errors can be identified by a high R-free (>40%) [Brünger, 1997].

To test the accuracy of a protein model, it is always important to inspect "omit" difference Fourier maps. To do this, one simply removes residues in the tested region from the model and refines the rest of the model until convergence. Usually, up to 10% of the model is omitted at a time. Omit electron density maps are then calculated using the equation:

$$\Delta\rho(xyz) = 1/V \sum_{h\,k\,l}\sum\sum |F_{obs}| - |F_{calc}| \exp[-2\pi i(hx+ky+lz) + i\alpha_{calc}].$$

The removed region does not contribute to the phase angle calculation because the phase angles α_{calc} and the structure factor F_{calc} are calculated using the rest of the model. This strategy produces an "unbiased" electron density map of the deleted region, and if the rest of the model is mostly correct, omit maps should suggest the correct position for the tested region. In addition, the location of solvent molecules is verified by inspection of the omit difference Fourier maps. This approach to model verification has proven to be very effective in revealing the location of errors (unlike R_{free}, which only informs you that such errors are present) and overbuilding of the protein model during the refinement.

Finally, one of the simplest and most sensitive means to assess the quality of a protein model is the Ramachandran plot. Due to steric hindrance, the main chain of a polypeptide usually assumes preferred, energetically favourable conformations. For each residue, these conformations are characterized by the value of two dihedral torsion angles, ϕ and ψ, which can only take a limited range of values. As an exception, glycines, lacking a side chain, occupy a far greater area of the conformational space. The distribution of ϕ and ψ is called Ramachandran plot. For all residues (but glycines), the ϕ and ψ values should fall within allowed values. If there is a considerable number of outliers, then the structure needs to be reassessed and improved. A good model should have most residues in the most favourable regions of the Ramachandran space and only a few residues in disallowed regions. However, sometimes, "disallowed" ϕ and ψ angles are observed for the residues involved in ligand binding [Gunasekaran et al., 1996; Lu et al., 1998].

Protein Science **6**: 2465-2468, 1997.

CHAPTER 3

PRELIMINARY STRUCTURAL STUDIES ON THE MULTI-LIGAND BINDING DOMAIN OF THE TRANSCRIPTION ACTIVATOR, BMRR, FROM *BACILLUS SUBTILIS*.

Ekaterina E. Zheleznova[1], Penelope N. Markham[2], Alexander A. Neyfakh[2] and Richard G. Brennan[1]*

[1]Department of Biochemistry and Molecular Biology
Oregon Health Sciences University
Portland, Oregon 97201-3098

[2]Department of Medicinal Chemistry and Pharmacognosy
University of Illinois at Chicago
Chicago, Illinois 60607

*Author to whom correspondence should be addressed.
Phone: (503)494-4427 FAX: (503)494-8393
E-mail: brennanr@ohsu.edu

Abstract

In the bacterium *Bacillus subtilis* the DNA-binding regulatory protein, BmrR, activates transcription from the multidrug transporter gene, *bmr,* after binding either rhodamine or tetraphenylphosphonium. These two compounds, which have no structural similarity, are also substrates for the bacterial multidrug transporter. BmrR belongs to MerR family of transcription activators, but is different from the other family members in its ability to bind unrelated small molecule activators. As an initial step in the elucidation of the mechanism by which BmrR recognizes rhodamine and tetraphenylphosphonium and activates transcription, we have crystallised the 125 amino acid residue carboxy-terminal dimerization/ligand-binding domain of the BmrR, named the BRC (BmrR C-terminus). Tetragonal crystals of ligand-free BRC take the space group $P4_12_12$, or its enantiomorph $P4_32_12$, with unit cell dimensions a = b = 76.3 Å, c = 96.0 Å, $\alpha = \beta = \gamma = 90^\circ$. Diffraction is observed to at least 2.7 Å resolution at room temperature. In addition, we determined the secondary structure content of ligand-free and rhodamine-bound BRC by circular dichroism. In the ligand-free form, BRC has considerable β-sheet content (41%) and little α-helix structure (13%). After BRC binds rhodamine, its β-sheet content increases to 47% while the α-helix structure decreases to 11%. The structure of BRC will provide insight not only into its multidrug recognition mechanism but could as well aid in the elucidation of the recognition and efflux mechanisms of Bmr and other bacterial multidrug transporters.

Keywords: multidrug recognition, transcription activator, BRC, crystallisation, X-ray diffraction, circular dichroism.

Introduction

Multidrug resistance is observed clinically as the unresponsiveness of organisms, both prokaryotes and eukaryotes, to a variety of chemical therapeutics. One of the major mechanisms by which multidrug resistance arises, is the active efflux of drugs and toxins via membrane transporters [Lewis, 1994; Gottesman & Pastan, 1993]. Such active transport results in the inability of drugs to affect their intracellular targets and thus, in the failure of the drug therapy. These transporters, which are called multidrug transporters (mdr) because of their remarkable ability to recognise and pump out structurally and chemically diverse toxic compounds, have gained considerable attention. Although studied extensively in eukaryotes [Gottesman & Pastan, 1993], mdrs are also common to bacteria [Paulsen et al., 1996; Lewis, 1994; Nikaido, 1994]. Yet in spite of intensive molecular and biochemical studies, the structural mechanisms by which multidrug transporters recognise their dissimilar substrates remain unclear [Gottesman & Pastan, 1993]. Undoubtedly, understanding this mechanism at the atomic level could open novel avenues to overcome drug resistance. However, these membrane-associated mdr transporters have yet to be purified in the amounts necessary for crystallographic studies.

In addition and as an alternative to the bacterial multidrug transporters, their transcription regulators represent potential targets for chemotherapeutic intervention. Furthermore, these cytosolic proteins can serve as readily available systems to elucidate the biochemical and structural mechanisms of multi-ligand recognition. One candidate for these studies is the 245 amino acid residue protein BmrR from *B. subtilis* . This dimeric protein activates transcription from the bacterial mdr transporter gene, *bmr,* [Neyfakh et al., 1991] after binding either rhodamine or tetraphenylphosphonium. These structurally dissimilar compounds are also substrates for the *B. subtilis* mdr transporter, Bmr [Ahmed et al., 1994]. Therefore, *B. subtilis* carries a two-fold defense system against multiple

toxic compounds in which not only the multidrug transporter but also its regulatory protein are mobilized to effect the removal of structurally diverse compounds (drugs) from the cell.

The amino terminal region of BmrR contains a putative helix-turn-helix motif and shows substantial sequence homology to the amino terminal DNA-binding domains of a family of bacterial regulatory proteins, which include SoxR from *Escherichia coli* [Gaudu & Weiss, 1996 and Hidalgo et al., 1997], the mercury resistance operon regulator, MerR, of several bacterial species [Summers, 1992; Ansari et al., 1995; Utschig et al., 1995] and $TipA_L$ from *Streptomyces lividans* [Holmes et al., 1993]. Whereas the carboxy terminal activation domains of the latter two proteins bind a single specific ligand, BmrR is different in its high-affinity binding of the structurally dissimilar molecules, rhodamine (K_d = 1.4 μM) and tetraphenylphosphonium ($K_d \approx 100$ μM) [Ahmed et al., 1994]. Recently, four additional cationic aromatic compounds have been identified as potential ligands [Markham et al., 1997]. The structural mechanisms by which members of this family bind their cognate promoters and small molecule activators is still poorly understood.

As an initial step in the elucidation of the stereochemical mechanism of multi-ligand binding by BmrR, a 125 amino acid residue carboxy terminal domain of BmrR (abbreviated as BRC), which is responsible for dimerisation and high affinity multi-ligand binding, has been crystallised. Furthermore, to investigate any ligand-induced changes in the secondary structure content of BRC, circular dichroism studies have been done on the ligand-free and rhodamine-bound BRC.

Results and Discussion

Tetragonal bipyramidal crystals of BRC typically grow to dimensions of 0.4 mm x 0.4 mm x 0.3 mm and take the space group $P4_12_12$ or its enantiomorph $P4_32_12$. The unit cell dimensions are: a = b = 76.3 Å, c = 96.0 Å and $\alpha = \beta = \gamma = 90°$. This crystal form

diffracts isotropically to at least 2.7 Å at room temperature. Intensity data have been collected on the ADSC multiwire area detector and processed with software provided by ADSC (Table 3.1).

Crystals of the selenomethionine-substituted BRC diffract to 3.0 Å at room temperature. An intensity data set on these crystals has been collected (Table 3.1). Difference Patterson analysis reveals three peaks in the asymmetric unit and indicates that the selenomethionine-incorporated BRC is a reasonable heavy atom derivative. As only three sites were found, it is very likely that there is a monomer in the asymmetric unit. Interestingly, a monomer per asymmetric unit results in a Matthews' coefficient (V_m) of 4.99 Da/Å^3, which is unusually high for proteins that diffract reasonably well [Matthews, 1968]. However, the assumption that there are two molecules per asymmetric unit yields a V_m of 2.49 Da/Å^3, which is more congruous with the diffraction quality of this crystal form. Thus, the number of protein molecules per asymmetric unit remains somewhat ambiguous.

Attempts to cocrystallise BRC with its ligands, rhodamine or tetraphenylphosphonium, were unsuccessful. However, we are able to introduce ligands into crystals via soaking for several hours in a crystallisation solution that contained 2 to 3 molar excess of these compounds. Overnight soaks resulted in crystal cracking.
The unusually high salt concentration (4.3 M NaCl) of the crystallisation solution prompted us to determine the oligomerization state of BRC. Using dynamic light scattering we found that in the presence of 0.1 M NaCl, a 2 mg/mL solution of BRC is monodisperse and has an estimated molecular weight of 45 kDa. Because the molecular weight of the monomer is 14 kDa, this result would indicate a trimeric BRC. However, if a dimeric protein has a nonspherical shape (for instance, elongated) its molecular weight might be overestimated. In addition, the existence of an elongated BRC dimer is consistent with previous analytical gel filtration studies that showed BRC migrates with an apparent molecular weight of 36 kDa which is greater than its calculated molecular weight of 28 kDa [Markham et al.,

1996]. Furthermore, gel-mobility shift assays and gel filtration chromatography studies on full-length BmrR-cognate operator complexes [Ahmed et al., 1994] also support the dimeric nature of BmrR and, consequently, of BRC. Additional dynamic light scattering experiments revealed that increasing the BRC concentration to 8 mg/mL in the 0.1 M NaCl solution leads to polydispersity and the formation of aggregates with estimated molecular weights from 300 to 1000 kDa. Interestingly, when the salt concentration is raised to 4 M NaCl, which approximates our crystallisation condition, the 8 mg/mL solution of BRC becomes monodisperse again. The molecular weight of the average particle has increased to 180 kDa, which can be rationalized as a tetrameric ensemble of BRC dimers.

The results from these dynamic light scattering experiments parallel our crystallisation results, in which crystallisation trials containing less than 3 mg/mL protein remain clear, while those made with protein solutions of 10 mg/mL form fluffy precipitates in the presence of 0.1 M to 0.5 M NaCl or produce three dimensional, data quality crystals when the NaCl concentration is raised to 4.3 M. Our observations support previous dynamic light scattering studies, which indicate that non discretely aggregating protein solutions do not crystallize readily, whereas monodisperse protein samples can yield diffraction quality crystals [Ferré-D'Amaré & Burley, 1994].

We also determined the secondary structure content of ligand-free and rhodamine-bound BRC by circular dichroism to investigate any ligand-induced changes in the secondary structure of the protein. The CD spectrum of the ligand-free BRC is shown in Figure 3.1. The variable selection method used to deconvolute this spectrum yielded a secondary structure of 13% helix, 41% β-sheet, 10% β-turn and 36% 'other' structural elements (Table 3.2). The high β-sheet and low α-helix composition was unexpected as secondary structure prediction methods overestimated the percentage of α-helix by as much as two-fold. To test whether binding of rhodamine induces a structural change in BRC we extended our analysis to determine the CD spectrum of BRC in the presence of rhodamine.

The CD spectrum of rhodamine alone was subtracted from the CD spectrum of BRC in the presence of rhodamine and the resulting difference CD spectrum of BRC in its ligand-bound form is shown in Figure 3.1. There is a distinct, though small, change in the CD spectrum of BRC upon binding of rhodamine, which, according to the variable selection method, represents a 6% increase in β-sheet content accompanied by a 2% decrease in the α-helix and a 5% decrease in the β-turn structure (Table 3.2). We assume that in the presence of BRC the CD signal of rhodamine does not change and therefore, conclude that the binding of rhodamine to BRC induces or stabilizes β-sheet structure. The crystal structure of the ligand-free and rhodamine-bound BRC will be necessary to confirm this. CD experiments with the BRC ligand, TPP, could not be done because of the strong UV absorption of the ligand at the saturating concentration.

Crystallographic studies on BRC will help unravel the stereochemical mechanism by which this protein binds the structurally different compounds, rhodamine and tetraphenylphosphonium. Furthermore, the BRC structure could possibly shed light on the substrate binding site of the multidrug transporter Bmr. The amino acid sequence identity between BRC and Bmr is approximately 26%, but the identical as well as homologous residues are scattered throughout the sequence. However, this clearly does not preclude the two proteins from having structurally homologous multi-ligand binding pockets. Such similarity might not be detected by a linear sequence comparison if the ligand-binding sites are formed by the three-dimensional arrangement of noncontiguous amino acid residues. Regardless, the crystal structure of the BRC in its unliganded form and in complex with rhodamine and tetraphenylphosphonium will reveal the atomic architecture of a ligand-binding pocket that favours the binding of multiple, structurally unrelated compounds.

Materials and Methods

Crystallisation and intensity data collection

The BRC, which encompasses amino acid residues 121-245 of BmrR, is prepared by trypsin cleavage of a BRC-thioredoxin fusion protein and purified as described previously [Markham et al., 1996]. BRC has been chosen for the studies as the intact BmrR protein is only sparingly soluble. The pure BRC is concentrated to 10-15 mg/mL in a solution of 20 mM tris·HCl, pH 7.5, and 0.1 M NaCl. Crystals of the apo BRC are grown at room temperature in hanging drops by the vapour diffusion method. The crystallisation solution is 4.3 M NaCl, 2 mM $MnCl_2$ and 20 mM tris·HCl, pH 8.5. The presence of Mn^{2+} ions is crucial for the production of diffraction quality crystals [Trakhanov & Quiocho, 1995]. Crystals take a month to appear and grow for two additional weeks. Rhodamine or tetraphenylphosphonium has been introduced into the crystals by soaking the native BRC crystals for several hours in a mother liquor, which contained 1.5 mM rhodamine or 2 mM tetraphenylphosphonium. X-ray intensity data have been collected at room temperature with an Area Detector Systems Corporation (ADSC) multiwire area detector using a Rigaku RU200-H rotating anode generator as the X-ray source (40 kV, 150 mA). The data were processed with software provided by ADSC [Xuong et al., 1985] and their relevant statistics are given in Table 3.1.

Preparation of the heavy atom derivatives

BRC contains three methionine residues and a selenomethionine-substituted BRC was prepared by transforming the BRC-thioredoxin-pBAce construct into *E. coli* DL41 cells (Yale Genetic Stock Center), a strain which is auxotrophic for methionine. The methionine normally included in the low phosphate induction media [Allen & Ullman,

1993] was replaced with selenomethionine at a final concentration of 40 mg/mL [LeMaster & Richards, 1985]. Cells were grown in the dark at 30°C for 24 hours. The selenomethionine-substituted protein was purified according to the previously reported protocol [Markham et al., 1996] except for the addition of 5mM DTT to prevent selenomethionine oxidation. The final yield of the selenomethionine-substituted BRC was 3 mg/L. Crystals of the selenomethionine-incorporated BRC grow under conditions described above and are isomorphous with the native crystals. An additional potential heavy metal derivative has been prepared by soaking native BRC crystals for 1 hour in 5 mM $KAuCl_4$. Relevant crystallographic data of the selenomethionine-substituted and gold-soaked crystals are given in Table 3.1.

Dynamic light scattering

Dynamic light scattering studies on BRC were done using a DynaPro-801 Dynamic Light Scattering Instrument (Protein Solutions). Protein samples at either 2 or 8 mg/mL were prepared in solutions of 20 mM tris·HCl buffer, pH 7.5, which contained either 0.1 M or 4.0 M NaCl. Prior to the experiment, all protein samples were filtered through 0.1 micron Anotop filters (Whatman) to eliminate any aggregated particles. All data were analyzed using AutoPro PC software. The reported values are averages of 10 scans of 30s each.

CD spectroscopy

CD spectra of BRC in the presence or absence of rhodamine were taken on a JASCO J-500A spectrometer. Measurements were made using a 0.1 mm path length cell (Helma), thermostatted at 20°C using a constant temperature bath (Lauda). The instrument was calibrated with solutions of (+)-10-camphorsulphonic acid ($\Delta\varepsilon = +2.37\ M^{-1}\ cm^{-1}$ at

290.5 nm and -4.95 M^{-1} cm^{-1} at 192.5 nm). Data were collected on an IBM/PC-XT using the IF-2 interface and software provided by Jasco. Spectra and buffer baselines were the averages of 10 scans, each recorded at 0.1 nm intervals with a scanning rate of 10 nm/min and a 2 s time constant. The protein concentration was determined by amino acid analysis and was approximately 1.3 mg/mL. The BRC samples subjected to the CD analysis contained either protein in 50 mM potassium phosphate buffer, pH 7.4, or the protein in 50 mM potassium phosphate buffer, pH 7.4., plus 0.1 mM rhodamine. Before spectral deconvolution for secondary structure analysis, the buffer baseline was subtracted and the spectra were smoothed using the smoothing software provided by Jasco. The protein spectrum in the presence of rhodamine is a difference spectrum. The CD spectrum of 0.1 mM rhodamine solution in 50 mM potassium phosphate buffer, pH 7.4, was subtracted from the spectrum of BRC measured in the presence of 0.1 mM rhodamine. To determine the secondary structure content the CD spectra were deconvoluted using the variable selection method described in detail in [Compton et al., 1987].

Acknowledgements

This research is supported by an N.L. Tartar Trust Fellowship from the Medical Research foundation of Oregon (E.E.Z.) and NIH grants GM49244 (R.G.B.) and GM49819 (A.A.N.). We thank Dr. H.P. Bächinger for his help with our circular dichroism studies.

Table 3.1. Summary of crystallographic data collected for ligand-free BRC and two heavy-atom derivatives.

	Native	SeMet	$KAuCl_4$
Resolution (Å)	2.7	3.0	3.0
Number of measurements	51,160	25,303	24,171
Number of reflections	8,272	5,692	5,887
Completeness (%)	100	96.0	100
R_{sym} (%)[1]	4.9	5.5	6.6
R_{iso} (%)[2]	-	9.6	23.4

[1] $R_{sym} = \Sigma|I_o - \langle I \rangle|/I_o$, where I_o is the observed intensity, and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry related reflections.

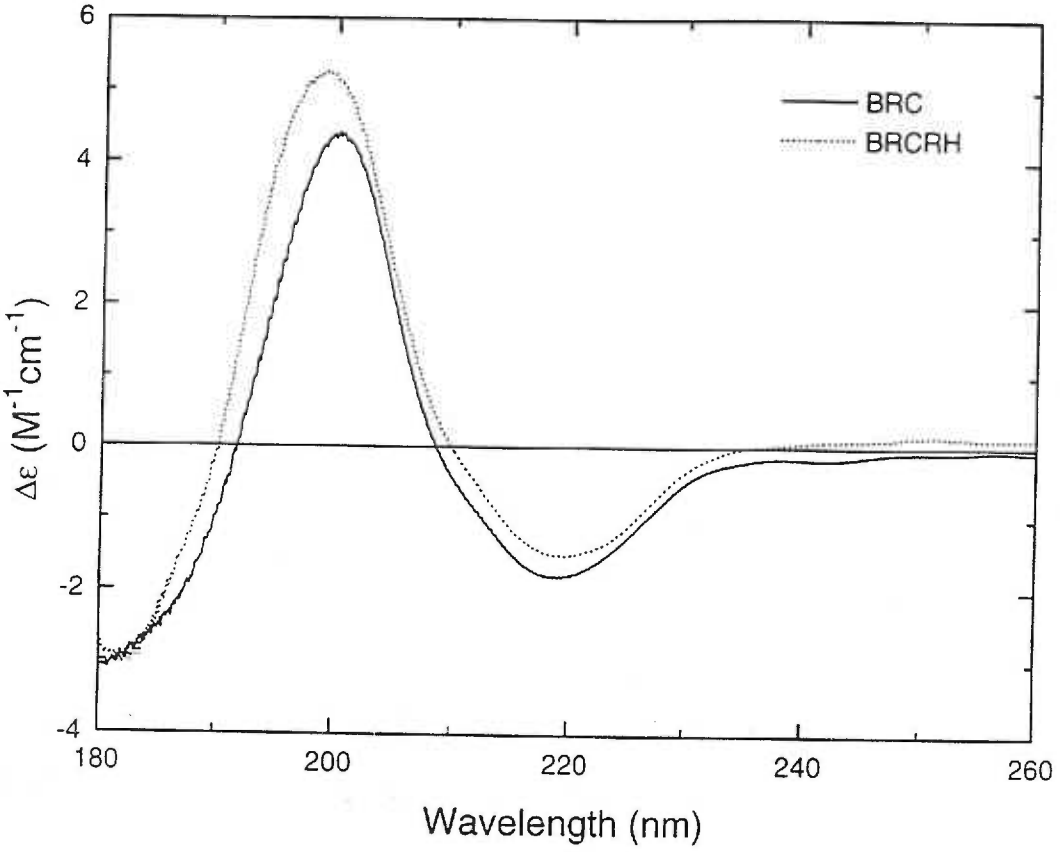
[2] $R_{iso} = \Sigma||F_{PH}|-|F_P|| / \Sigma|F_P|$, where $|F_P|$ is the protein structure factor amplitude, and $|F_{PH}|$ is the heavy atom derivative structure factor amplitude.

Table 3.2. Secondary structure content of ligand-free and rhodamine-bound BRC.

	H	B	T	O	Total
BRC	0.13	0.41	0.11	0.36	1.01
BRC-RH	0.11	0.47	0.06	0.36	1.00

The secondary structure contents determined by the variable selection method for the spectra shown in Figure 3.1. The abbreviations used are: **H**, α-helix; **B**, β-sheet; **T**, β-turn; **O**, other. BRC corresponds to ligand-free protein and BRC-RH is BRC bound to rhodamine.

Figure 3.1. CD spectra of ligand-free and rhodamine-bound BRC. The change in the molar amino acid ellipticity, $\Delta\varepsilon$, ($M^{-1}\cdot cm^{-1}$, ordinate), is plotted against the wavelength, (nm, abscissa). The solid line (—) represents the spectrum of ligand-free BRC (**BRC**) in 50 mM potassium phosphate buffer, pH 7.4, and the dashed line (---) represents the spectrum of rhodamine-bound BRC (**BRC-RH**) in the same buffer. The concentration of BRC used for both spectra is approximately 1.3 mg/mL and the concentration of rhodamine in the **BRC-RH** sample is 0.1 mM. All spectra were recorded at 20° C.


BRC
BRCRH
Δε (M-1cm-1)
Wavelength (nm)
6
4
2
0
-2
-4
180
200
220
240
260

Cell **96**: 363-362, 1999.

CHAPTER 4

STRUCTURAL BASIS OF MULTIDRUG RECOGNITION BY BMRR, A TRANSCRIPTION ACTIVATOR OF A MULTIDRUG TRANSPORTER.

Ekaterina E. Zheleznova*, Penelope N. Markham†‡, Alexander A. Neyfakh† and

Richard G. Brennan*§

*Department of Biochemistry and Molecular Biology,

Oregon Health Sciences University, Portland, Oregon 97201-3098

† Center for Pharmaceutical Biotechnology,

University of Illinois at Chicago, Chicago, Illinois 60607

‡Present address: Influx Inc., Chicago Technology Park, Chicago, IL 60612, USA.

§Author to whom correspondence should be addressed.

Phone: (503) 494-4427 FAX: (503) 494-8393

E-mail: brennanr@ohsu.edu

Abstract

Multidrug-efflux transporters demonstrate an unusual ability to recognize multiple structurally dissimilar toxins. A comparable ability to bind diverse hydrophobic cationic drugs is characteristic of the *Bacillus subtilis* transcription regulator BmrR, which upon drug binding activates expression of the multidrug transporter Bmr. Crystal structures of the multidrug-binding domain of BmrR (2.7 Å resolution) and of its complex with the drug, tetraphenylphosphonium (2.8 Å resolution), revealed a drug-induced unfolding and relocation of an α helix, which exposes an internal drug-binding pocket. Tetraphenylphosphonium binding is mediated by stacking and van der Waals contacts with multiple hydrophobic residues of the pocket and by an electrostatic interaction between the positively charged drug and a buried glutamate residue, which is the key to cation selectivity. Similar binding principles may be used by other multidrug-binding proteins.

Introduction

One of the major mechanisms underlying multidrug resistance in both prokaryotes (Nikaido, 1994; Paulsen et al., 1996) and eukaryotes (Gottesman and Pastan, 1993; Sharom, 1997; Goffeau et al., 1997) is the active extrusion of numerous structurally unrelated cytotoxic compounds by membrane proteins known as multidrug-efflux transporters. The broad substrate specificity displayed by these transporters contrasts dramatically with the narrow chemical specificity of the vast majority of ligand-binding proteins. Despite extensive biochemical studies, the structural mechanisms which multidrug transporters use to recognize dissimilar compounds remain obscure, primarily because of the dearth of structural information on any of these membrane proteins.

The phenomenon of multidrug recognition is not confined to multidrug transporters. In bacteria, several transcription regulators of these transporters have been demonstrated to promote transporter expression in response to structurally dissimilar toxic compounds (Lomovskaya et al., 1995; Ma et al., 1996, Grkovic et al., 1998). For one such regulator, the *Bacillus subtilis* protein BmrR (Ahmed et al., 1994), which controls the transcription of the multidrug transporter gene *bmr* (Neyfakh et al., 1991), the binding of diverse inducers has been demonstrated directly. Structurally dissimilar substrates of the Bmr transporter, such as tetraphenylphosphonium (TPP) and rhodamine (Ahmed et al., 1994), as well as a number of other aromatic cationic compounds (Markham et al., 1997), have been shown to bind to BmrR and stimulate transcription from the *bmr* promoter. Unlike multidrug transporters, the cytosolic BmrR is amenable to structural studies and thus, provides a promising model system for understanding the structural principles of multidrug recognition.

BmrR belongs to the MerR family of transcription regulators (Sadowsky et al., 1991; Summers, 1992; Holmes et al., 1993; Ansari et al., 1995; Utschig et al., 1995; Gaudu and Weiss, 1996; Hidalgo et al., 1997). This protein is composed of an amino-terminal helix-turn-helix DNA-binding domain (residues 1-90) homologous to the DNA-binding domains of other MerR-type proteins (Ahmed et al., 1994), a linker region (residues 91-120), and a carboxy-terminal domain (residues 121-279). The latter domain, named BRC (BmrR C-terminus), has no sequence homology to other proteins of the MerR family or any other protein in current databases. Purified BRC forms dimers and binds the aromatic cationic ligands rhodamine and TPP with the same affinity as the full-length BmrR: $K_{d,Rho}$ = 1.5 μM and $K_{d,TPP}$ = 100 μM (Markham et al., 1996). To elucidate the structural and stereochemical mechanisms of multidrug recognition by BmrR, we undertook crystallographic studies on its carboxy-terminal drug-binding/dimerization domain. Here, we report the structures of drug-free BRC to 2.7 Å resolution and TPP-bound BRC to 2.8 Å resolution.

Results and Discussion

Apo BRC Structure Determination

The structure of the apo BRC was solved by multiple isomorphous replacement (Blundell and Johnson, 1976; Drenth, 1994). The R_{factor} for the final model is 20.2% (Table 1). It includes 143 residues, 30 water molecules and one Mn^{2+} ion. Residues 42-47, 77-80, and 142-147 are disordered and not included in the final model (although BRC

constitutes residues 121-279 of BmrR, in this paper, we refer to them, for convenience, as residues 1-159). The stereochemical quality of the model is excellent with no Ramachandran outliers, as assessed by the program PROCHECK (Laskowski et al., 1993).

Overall Structure of the apo BRC

The BRC monomer is composed of eight β strands and three α helices (Figure 4.1A). The core of the protein is created primarily by the β strands, which form an antiparallel β sheet with the topology β1–β6–β8–β7–β3–β4–β2–β5. BRC has a unique fold as determined by the three-dimensional homology-search algorithm Dali (Holm and Sander, 1993), but its structure resembles the β barrel motif taken by members of the lipocalin family (Sacchettini and Gordon, 1993; LaLonde et al., 1994; Flower, 1996). One side of the β sheet core is covered by a 17-residue α helix, α3 (residues 107-123); a well-defined 12-residue loop (residues 57-68); a 2-turn α helix, α1 (residues 25-30); and a 3-turn α helix, α2 (residues 34-42), which are connected by a short turn. Surprisingly, the crystal structure revealed the proteolytic cleavage of the C-terminal β8 strand (residues 148-158) from the preceding β7 strand which occurs during the trypsin digestion step in the course of protein purification. Protein sequencing and electrospray mass-spectrometry confirmed that trypsin cleaves BRC between residues Lys144 and Lys145. However, in the crystal structure, β8 strand is integrated into the core β sheet and remains associated with the protein.

Structure of the Dimer

There is one monomer in the asymmetric unit of the crystal. However, biochemical data obtained from dynamic light scattering and gel-filtration experiments indicate that BRC is dimeric (Markham et al., 1996; Zheleznova et al., 1997). Moreover, the full-length BmrR binds its operator as a dimer (Ahmed et al., 1994). To identify the biologically relevant dimer, we considered the three potential dimers that are created by the symmetry operators of the tetragonal unit cell. In one of them, both subunits are related by a crystallographic 2-fold symmetry (Figure 4.1B). Furthermore, the buried surface area of this dimer is ~900 $Å^2$ per monomer (Honig and Nicholls, 1995), whereas each of the two alternative dimers buries only ~400 $Å^2$, which strongly indicates that the symmetrical dimer corresponds to the biologically relevant one. Nearly 65% of the buried dimer interface is hydrophilic, and 8 ionic and hydrogen-bonding interactions stabilize the dimer. These involve OE1 of Glu12 and NH1 of Arg156; OE2 of Glu12 and NH2 of Arg156; NE2 of Gln58 and the carbonyl oxygen of Ile84; NE1 of Arg69 and the carbonyl oxygen of Met88; and their symmetry-related mates.

The most unexpected feature of the dimer interface is the presence of a subunit-tethering divalent metal ion residing on a crystallographic 2-fold axis. The metal is tetrahedrally coordinated by the side chains of Glu19 (Me^{2+}–OE2, 1.8 Å) and His70 (Me^{2+}–NE2, 1.9 Å) and their two-fold symmetry mates, Glu19' and His70' (Figure 4.1B). This metal ion has been refined as a Mn^{2+} because of the presence of 2 mM $MnCl_2$ in the crystallization solution, but the relevant *in vivo* ion is unknown. The importance of the divalent metal ion in dimer stability was indicated by two site-directed

mutants His70Ala and Glu19Gln. The His70Ala mutant, which lacks two of the four metal-coordinating residues, is completely insoluble, whereas the Glu19Gln mutant, in which the metal binding should be weaker, retains solubility but, unlike the wild-type variant, precipitates upon storage.

Internal binding site

The apo BRC structure reveals no pocket or cavity that would appear to serve as an obvious drug-binding site. However, examination of the hydrophobic core of the protein revealed an anomaly, a completely buried acidic residue, Glu134. The burial of a charged residue in the protein core is normally considered energetically unfavorable, yet in this structure, the carboxylate group of the glutamate is "neutralized" by hydrogen bonds to the hydroxyl groups of Tyr33 from the N-cap of helix α2, Tyr68 from the loop connecting strands β3 and β4, and Tyr110 from helix α3 (Figure 4.2). Such an arrangement, a negatively charged side chain surrounded by aromatic moieties, would make an attractive binding site for aromatic/hydrophobic cationic drugs. Consequently, the putative role of Glu134 in drug binding was examined by substituting this residue with glutamine, alanine or lysine and by assaying rhodamine binding to each mutant (Markham et al., 1996). The stability and solution properties of all three mutant proteins were indistinguishable from those of the wild-type BRC, indicating that the mutant proteins were folded like the wild-type protein. In addition, the Glu134Gln mutant crystallized under the same conditions and in the same space group as the wild-type BRC. None of the mutants, however, was able to bind rhodamine to any measurable extent (at least 20-fold reduction as compared to wild type), strongly supporting the key

electrostatic role of residue Glu134 in drug binding. It remained unclear, however, how Glu134 could participate in drug binding considering that it is completely buried within the hydrophobic core of the protein.

BRC-TPP Complex Structure Determination

To determine the detailed architecture of the multidrug-binding site of BRC and to gain insight into the mechanism of drug binding, we crystallized BRC complexed to one of the BmrR inducers and BRC ligands, tetraphenylphosphonium (TPP). The structure of the BRC-TPP complex was solved with the molecular replacement program EPMR (Kissinger, unpublished) using the apo BRC structure as a search model. There is one monomer in the asymmetric unit. The current model of the BRC-TPP complex to 2.8 Å resolution has an R-factor of 23.5% and includes 130 residues, 11 water molecules, one TPP molecule, one divalent metal ion, and one sulfate ion (Table 1). Residues 1, 34-49, 78-80 and 141-148 are disordered and are not included in the final BRC model.

Overall Structure of the BRC-TPP Complex

The global structures of the TPP-bound BRC monomer and dimer are very similar to those of the apo protein (Figure 4.3). An overlay of the Cα atoms of residues 2-27, 50-76, 81-140, and 149-158 of the apo and drug-bound monomers results in a root-mean-square difference (RMSD) of 0.81 Å. There are no gross differences in the dimerization interface of BRC in response to TPP binding, and the RMSD for the apo and drug-bound dimers is 0.83 Å after superimposition of the Cα atoms of residues 2-27, 50-76, 81-140, and 149-158 from both subunits. However, when the flexible loop residues 76 and 81-83,

which neighbour disordered residues 77-80, are excluded from the overlay, the RMSD is only 0.61 Å for the monomer and 0.63 Å for the dimer. As found in the dimer interface of the apo protein, a divalent metal ion is coordinated identically by residues Glu19, His70, and their dyadic mates. As no metal salts were added to the crystallization solution, the metal ion appears to be carried from the cell through protein purification and crystallization, and its identity is unknown.

TPP binding pocket

The TPP binding pocket is created by residues which originate from helix α1 (Ile23, Val28) and strands β3 (Tyr51, Ala53), β4 (Tyr68, Ile71) and β7 (Ile136). The TPP molecule wedges into the binding cavity using one of its four phenyl rings as a lead prong (Figure 4.4). Each of the four rings of the TPP molecule contributes to binding and makes van der Waals contacts to the side chains of the above five hydrophobic residues while one phenyl ring stacks against the aromatic side chain of residue Tyr51 and another phenyl ring stacks against the aromatic side chain of residue Tyr68. As anticipated from the structure of the apo protein and mutational studies, the key electrostatic component of TPP binding is residue Glu134, which is located at the bottom of the binding pocket. The negatively charged carboxylate group of Glu134 makes an electrostatic contact with the positively charged TPP molecule. In the TPP molecule, the formal positive charge is carried by the phosphorus atom of TPP, which is 7.8 Å away from the carboxylate oxygen OE2. However, since TPP is a conjugated molecule, its phenyl rings can carry a partial positive charge. The distance between the carboxylate oxygen OE2 and the closest

atom of TPP is 3.2 Å, and the low dielectric constant (~2) of the protein interior would strongly favor this electrostatic interaction. As seen from the distribution of the electrostatic surface potential, Glu134 creates the negative charge inside the binding pocket (Figure 4.5, right) necessary to attract and bind a positively charged drug, which is consistent with the inability of the Glu134 mutants to bind rhodamine. Finally, to complement its electrostatic potential, the carboxylate oxygens of Glu134 maintain their hydrogen bonds to nearby residues Tyr110 and Tyr68.

Conformational changes

TPP binding induces dramatic local conformational changes in the BRC protein, which make the binding pocket accessible to the drug. These changes involve helices α1 and α2 and the turn connecting them (Figure 4.3, bottom). In contrast to the apo structure, helix α1 is rotated ~10° around its amino-terminus, and its carboxy-terminus (residue Asn30) is unwound. The tight turn connecting helices α1 and α2, residues Ala31-Ser32-Tyr33, unfolds and extends into the solvent such that the Cα atom of Tyr33 is translated 11.5 Å from its apo position and its side chain is completely removed from the protein core. The repositioning of residue Tyr33 is especially important because in the apo structure, Tyr33 contributes one of the internal hydrogen bonds to the buried carboxylate of Glu134 (Figure 4.2), a bond that is broken when drug binding takes place. A water molecule, Wat1, occupies the void created by the displaced Tyr33 side chain and forms a hydrogen bond to Glu134 (Figure 4.4).

The most important structural change in the BRC-drug complex involves helix α2 (residues 34-42), which is now disordered. The disordering extends to residue 49, including the flexible-loop residues 43-48, which are already disordered in apo BRC. Increased flexibility or disordering of helix α2 was suggested from protease cleavage studies (not shown) which demonstrated that rhodamine or TPP binding dramatically increase trypsin cleavage between residues Lys37 and Lys38. The observed unfolding of helix α2 is critical for drug binding. In the absence of drugs, helix α2 lies across the drug-binding pocket and completely blocks access to the binding site (Figure 4.3, top; Figure 4.5, left). Furthermore, helix α2 stabilizes the apo conformation by inserting residues Tyr33 and Leu36 into the core of the protein with the former residue hydrogen bonding to Glu134 (Figure 4.2) and the latter making van der Waals contacts to core residues Val28 and Ile23. To bind TPP, helix α2 must relocate which, in turn, requires the helix to unfold because of the stereochemical constraints of the BRC structure, i.e., a folded 9-residue α-helix would be too short to connect residue Ser34 to residue Ser42 (Figure 4.3, bottom). These drug-dependent conformational changes lead to the exposure of the internal drug-binding pocket where the phenyl moieties and charge of TPP sterically and functionally replace the side chains of Tyr33 and Leu36 and thereby stabilize the drug-bound conformation.

The ligand-induced helix to coil transition undergone by BRC is uncommon and contrasts with the reverse process that is observed in several DNA-binding proteins and the activation domains of certain transcription factors upon ligand binding (O'Neil et al., 1991; Spolar and Record, 1994; Ferre-D'Amare et al., 1994; Schumacher et al., 1995;

Nagadoi et al., 1995; Kussie et al., 1996; Spronk et al., 1996; Uesugi et al., 1997). However, the DNA-binding proteins, restriction enzyme BamHI (Newman et al., 1995) and transcription factor Ets-1 (Petersen et al., 1995; Donaldson et al., 1996), show similar helical unfolding. When BamHI binds its specific DNA site, the C-terminal helix unfolds into an extended arm that makes additional contacts to its DNA ligand. More analogous to BRC, specific DNA binding by Ets-1 induces unfolding of an α-helix in the DNA-binding inhibitory region of the protein. This helix blocks access of the recognition helix to its cognate DNA site and must unfold to allow high affinity binding.

Mechanism of ligand entry

The structures of apo and TPP-bound BRC capture the end states of the drug-binding event. A key question is how does TPP gain entrance to the internal binding site of BRC (BmrR). The electrostatic surface potential surrounding the binding site (Figure 4.5, left) suggests a pathway that involves residue Glu21. This residue is proximal to helix α2 and creates a negative patch on the protein that could initially attract and possibly bind TPP and other cationic ligands of BRC. From here, the positively charged molecules could slide toward the drug-binding pocket. However, drug binding cannot occur unless helix α2 unfolds and moves out of the way. As intimated above, this 9-residue α helix is likely to possess an intrinsic flexibility and to fluctuate between a folded and partially unfolded conformation, thus allowing drugs to enter the internal binding site. Indeed, the average thermal parameter of the Cα atoms of helix α2 is 52Å^2, which is

significantly higher than the average thermal parameter (32 $Å^2$) of the Cα atoms of the remainder of the protein. Similar flexibility and its implied importance in ligand binding has been observed for several members of the lipocalin family (Chen et al., 1998; Constantine et al., 1998). Finally, the aromatic cationic drugs themselves might actively destabilize helix α2 by repelling the positively charged lysines Lys35, Lys37, and Lys38 located on the outer face of this helix.

Model of the BRC-rhodamine complex

Our attempts to obtain data-quality crystals of BRC complexed to other drugs, rhodamine in particular, have been unsuccessful. However, using the TPP-BRC structure as a template, a rhodamine-BRC complex was modelled by manually docking rhodamine in the binding pocket. Subsequently, the DOCK option in the software package Sybyl (Tripos) was used to minimize the bad contacts between the two molecules by allowing the position of the rhodamine molecule to change whilst the protein structure was held fixed. From this BRC-drug complex model (Figure 4.6), it is clear that rhodamine is not only able to make van der Waals and stacking contacts similar to those observed in the TPP-BRC complex, but can also penetrate deeper into the core of BRC. Specifically, the positively charged nitrogen of the amino ethyl group of rhodamine can be positioned as close as 2.9 Å from the oxygen of the carboxylate group of Glu134, in contrast to the 7.8 Å distance for the positively charged phosphorus of TPP, without steric clash. The closer approach of the positive charge on rhodamine to the carboxylate of Glu134 likely accounts for most of its ~100-fold higher binding affinity in comparison with TPP. The

modelled rhodamine-BRC structure, together with TPP-BRC complex, emphasizes the critical role of the steric complementarity in discrimination against other, less planar or more branched, hydrophobic cations. Nevertheless, the precise description of the rhodamine-BRC and other drug-BRC complexes awaits the determination of their crystal structures.

Transcription activation

Binding of drugs to the BRC domain converts BmrR into an activator of transcription from the *bmr* promoter. This activation, similar to other MerR-type proteins (Summers, 1992; Holmes et al., 1993; Ansari et al., 1995; Gaudu and Weiss, 1996; Hidalgo et al., 1997), is likely to occur through untwisting of the spacer region of the promoter, which serves as the BmrR-binding site (Ahmed et al., 1994). This untwisting leads to the proper positioning of the promoter motifs binding RNA polymerase and thus initiates transcription (Summers, 1992; Ansari et al., 1995; Hidalgo et al., 1997). The structures of the apo BRC and the TPP-bound protein reveal no large conformational or structural changes outside of the drug-binding region (Figure 4.3). Therefore, it can be hypothesized that the unwinding and relocation of helix α2 serves as an inducing signal between the drug-binding and the DNA-binding domains of BmrR. For example, the DNA-untwisting activity of the N-terminal domain of BmrR can be repressed by its intra or inter-subunit contacts with helix α2. Through the disordering of helix α2, drug binding would impair such an interaction and relieve repression. The 30-residue linker, connecting the DNA-binding and the drug-binding domains is more than

sufficient to allow this possibility. Clearly, the determination of the molecular mechanism by which drug binding turns BmrR into a transcription activator entails further biochemical investigation and, ultimately, the structure of a BmrR-promoter-drug ternary complex.

In summary, the x-ray structures of apo BRC and the TPP-BRC complex reveal the helix-to-coil transition of helix α2 that is required for high-affinity drug binding. Moreover, the structures identify residue Glu134, which is completely solvent inaccessible in the apo structure, as the key to cation selectivity, while a number of nonpolar and aromatic side chains impose specific requirements on drug size and shape. The interactions between BRC and tetraphenylphosphonium suggest a similar binding mode for other hydrophobic cationic ligands of BmrR. It is tempting to speculate that other multidrug-binding proteins, including multidrug transporters, utilize some of these stereochemical and structural principles, such as relocation or disordering of secondary structure elements to expose otherwise inaccessible key residues, to bind their structurally dissimilar ligands and substrates.

Experimental procedures

Crystallization

Crystals of the 158-residue drug-free BRC (apo BRC), which take the tetragonal space group $P4_12_12$ with unit cell dimensions: a = b = 76.3 Å, c = 96.0 Å, $\alpha = \beta = \gamma = 90^{\circ}$, crystallized out of solutions containing 4.3 M NaCl, 20 mM tris·HCl, pH 8.5, and 2 mM $MnCl_2$, as described previously (Zheleznova et al., 1997). Crystals of BRC-TPP complex were grown from solutions containing 10 mM TPP, 1.5 M $(NH_4)_2SO_4$, 0.1 M sodium cacodylate, pH 6.5, and 5% 2-methyl-2,4-pentanediol. These crystals assume the hexagonal space group $P6_22$ with unit cell dimensions: a = b = 82.6 Å, c = 104.5 Å, $\alpha = \beta = 90^{\circ}$, $\gamma = 120^{\circ}$.

X-Ray Data Collection and Processing

X-ray intensity data for apo BRC were collected to 2.7 Å resolution at room temperature with an Area Detector Systems Corporation (ADSC) multiwire area detector (Xuong et al., 1985) using a Rigaku RU200-H rotating anode X-ray generator with a graphite monochromator (40 kV, 150 mA). The data were processed with software provided by ADSC. X-ray intensity data for the BRC-TPP complex were collected to 2.8 Å resolution at room temperature on an R-AXIS IV imaging plate system using a RIGAKU RU300 rotating anode X-ray generator equipped with double focusing mirrors and operating at 50 kV and 100 mA. The data were processed using BIOTEX (Molecular Structure Corporation, Inc., Woodlands, TX).

Structure determination of apo BRC

The structure of apo BRC was solved by multiple isomorphous replacement (MIR) (Blundell and Johnson, 1976; Drenth, 1994). Heavy atom derivatives were prepared by soaking native crystals in solutions of $KAuCl_4$ or $HgCl_2$, and the selenomethionine-substituted protein (LeMaster and Richards, 1985) was also a useful derivative (Table 1). Relevant intensity data collection statistics are listed in Table 1. Initially, one heavy atom site was identified in the difference Patterson map of the gold derivative. Difference and cross-difference Fourier analyses allowed the identification of the second gold site, one mercury site and one selenium site. The heavy atom parameters were refined via maximum likelihood to 3.0 Å resolution using the software package PHASES-95 (Furey and Swaminathan, 1997). At this stage, we were able to resolve the space group ambiguity between the enantiomorphs $P4_12_12$ and $P4_32_12$ in favor of the former. The overall figure of merit was 0.45 in the $P4_12_12$ space group while in the $P4_32_12$ space group it was only 0.35. The initial MIR map was uninterpretable, and solvent flattening (70% estimated solvent) as implemented in PHASES-95 was used. The solvent flattened map was readily interpretable, and had the correct handedness as revealed by the presence of a good electron density for a right-handed helix. Besides this α-helix, 7 β-strands were also obvious, and a polyalanine trace was fit to the density. Iterative cycles of model building in O (Jones et al., 1991) and phase combination, as implemented in PHASES-95, resulted in the addition of two more α-helices and one β-strand. This model included residues 1-42, 48-76, 81-141, and 148-158. The initial R_{factor} of the model was 42.6%.

Structure determination of BRC-TPP Complex

The structure of the BRC-TPP complex was solved by molecular replacement with the program EPMR, which finds crystallographic molecular replacement solutions using an evolutionary search algorithm (C.R. Kissinger and D.K. Gehlhaar, unpublished). The program optimizes the three rotational and three positional parameters for the search model with respect to a calculated correlation coefficient between F_O and F_C. At the end of this evolutionary optimization, a rigid-body refinement of the orientation and position of the search model is performed. We used data in the resolution range 15.0 Å to 4.0 Å and the entire apo BRC monomer as a search model. Both $P6_222$ and enantiomorph $P6_422$ were tested, but only $P6_222$ yielded a correct solution. $P6_422$ did not yield a solution. The correct solution had a correlation coefficient of 65.1% and a R_{factor} of 37.4% (Table 1).

Model refinement: apo BRC and BRC-TPP Complex

Prior to refinement, 10% of all the data were set aside for cross-validation purposes. The initial apo BRC model underwent positional least-squares refinement using data in the resolution range 20.0 Å to 3.0 Å, as implemented in software package TNT (Tronrud et al., 1987; Tronrud, 1997). After convergence, the R_{factor} was 27.8%. The resolution was extended to 2.7 Å, and positional and tightly restrained thermal-parameter refinement was carried out. Intervening model rebuilding was done using O. At this stage, solvent molecules were positioned into overlapping *2Fo-Fc* (>1σ) and *Fo-Fc* (>3σ) densities that were located within 3.4 Å distance from a hydrogen donor or acceptor. Their presence was additionally verified with omit maps in which these

molecules were removed, the model refined and *Fo-Fc* omit electron density maps calculated and inspected. Only solvent molecules with B-factor less than 80 were included in the final model. In the apo structure, the subunit-tethering metal ion was refined as Mn^{2+} at the occupancy of 0.5 because of its location on a crystallographic two-fold axis (Tronrud, 1997). The final model was verified by a series of *Fo-Fc* omit electron density maps in which 10% of the model was omitted and the remaining structure refined until convergence (typically 50 cycles). This approach to model verification has proven to be very effective in revealing errors and overbuilding of the protein model. The final R_{free} was 27.4%. The model was refined against all the data, which resulted in an R factor of 20.2%. The final protein model includes residues 1-42, 48-76, 81-141, and 148-158 and displays an excellent stereochemistry with no outliers. Residues Arg1 was refined as an alanine because the density for its side chain is missing. Relevant crystallographic statistics are listed in Table 1.

In the refinement of the BRC-TPP complex, coordinates from the molecular replacement solution were used as the initial model. This model underwent rigid-body refinement using data from 16.0 Å to 3.5 Å resolution, which dropped the R_{factor} from 37.4% to 35.0%. Positional refinement followed using data in the resolution range 16.0 Å to 3.0 Å. At this point, *2Fo-Fc* and *Fo-Fc* electron density maps were calculated and inspected. No density was found for helix α2 in either map. Instead, a tetrahedrally shaped density, indicative of a TPP molecule, was found at this location. To ascertain that the latter density we found was not a residual density from helix α2, residues 31-42 were deleted, and the initial rigid-body refinement was carried out again from 16.0 Å to

3.5 Å. The R_{factor} dropped to 29.0%. Positional refinement followed and was carried out to convergence with intensity data in the resolution range 16.0 Å to 3.0 Å. At this stage it was clear that helix α2 was completely disordered, as no continuous electron density was seen even at very low contour levels. However, the tetrahedrally shaped electron density feature remained, and a TPP molecule (Flomer et al., 1996) was fit into it. Subsequently, the BRC-TPP model underwent rounds of positional refinement from 16.0 Å to 3.0 Å and model rebuilding in O. The refinement converged, and the R_{factor} was 25.0%. At this stage, the resolution was extended to 2.8 Å, and tightly restrained B-factor refinement was carried out. After convergence, solvent molecules were positioned into the overlapping *2Fo-Fc* and *Fo-Fc* densities found within 3.4 Å distance from a hydrogen donor or acceptor. A sulfate ion was built into a large electron density feature located between two Arg side chains. Locations of waters and sulfate were verified by their removal followed by refinement until convergence and inspection of resulting *Fo-Fc* omit electron density maps. Only solvent molecules with B-factors less than 80 were included in the final model. As in the apo structure, the subunit-tethering metal ion was refined as Mn^{2+} ion at the occupancy of 0.5. The model was cross-validated by monitoring R_{free} calculated using 10% of the data, which was excluded from refinement. In addition, *Fo-Fc* omit electron density maps were calculated during all stages of the refinement. The final R_{free} is 32.6%, which reflects the limited resolution of the data that compromises the refinement of the structure and increases the difference between R_{free} and R_{factor} (Brünger, 1997). However, this R_{free} is well within the suggested range (Brünger, 1997). At the final stage, the model was refined against all the data yielding a final R_{factor} of 23.5%. The final protein model contains residues 2-33, 50-77, 81-140, and

149-158. Residues Tyr33, Thr75, Ser82, and Ser83, were refined as alanines because the density for their side chains is completely missing. The stereochemical quality of the structure is very good (Laskowski et al., 1993) (Table 1), but there are two residues Ser62 and Ser83 that are outliers on the Ramachandran plot. Residue Ser83 is located in a flexible loop which neighbours a disordered region, and residue Ser62 is located in a turn region; their electron densities are definable but poor. Relevant crystallographic statistics is listed in Table 1.

Coordinates for the apo BRC (1BOW) and TPP-bound BRC complex (2BOW) have been deposited in the Brookhaven Protein DataBase.

Mutagenesis

Each mutant variant of BRC was obtained by cloning an appropriately mutagenized PCR product encompassing the BRC-encoding region into the pHPThioFus expression vector, followed by protein expression and purification as described for the wild-type BRC (Markham et al., 1996). Mutagenized PCR products were obtained by crossover PCR with two standard primers defining the ends of the PCR product and two complementary mutagenic primers. Each mutant construct was resequenced to verify the correctness of the substitution and the absence of spontaneous PCR-generated mutations.

Acknowledgements

This research was supported by an N.L.Tartar Trust Fellowship from the Medical Research Foundation of Oregon to E.E.Z. and by NIH grants (GM49819) to A.A.N. and (GM49244) to R.G.B.

Table 4.1. Crystallographic data.

	Apo BRC	SeMet	$KAuCl_4$	$HgCl_2$	TPP complex
Resolution (Å)	24.0-2.7	10.0-3.0	10.0-3.0	10.0-3.0	42.2-2.8
Total observations	65,480	25,303	24,171	25,560	24,335
Unique reflections	8,264	5,692	5,887	3,036	5,322
Completeness (%)	100	96	100	51	95
Overall $I/\sigma(I)$	10.3	10.2	11.0	11.1	14.7
R_{sym} (%)[1]	6.0	5.5	6.6	5.3	8.3
Highest resolution shell[2]:					
Completeness (%)	100				90
$I/\sigma(I)$	1.6				3.8
Multiple isomorphous replacement - apo BRC					
R_{iso} (%)[3]		9.6	23.4	12.4	
Number of sites		1	2	1	
R_{cullis}[4]		0.49	0.69	0.57	
Phasing power[5]		1.0	1.7	1.5	
Overall figure of merit[6]	0.45 to 3 Å				
Molecular replacement - TPP complex					
Correlation coefficient (%)					65.1
Initial R_{factor}[7] (%)					37.4
Refinement -	**apo BRC**				**TPP complex**
Resolution range (Å)	20.0-2.7				16.0-2.8
Reflections used in refinement	8,250				5,285
Completeness, %	100				95
Amino acid residues	143				130
Protein atoms	1,171				1,045
Solvent molecules	30				11
Sulfate molecules					1
Mn^{2+} atoms	1				1
R_{factor}[7] (%)	20.2				23.5
R_{free} (%)	27.4				32.6

Table 4.1. (cntd).

RMS deviations		
Bond length (Å)	0.011	0.014
Bond angles (°)	1.29	1.72
B-factor ($Å^2$)	4.70	4.03
PROCHECK: Residues in Ramachandran regions (%, number of residues)		
most favourable	82.9 (102)	74.5 (82)
allowed	14.6 (18)	20.9 (23)
generously allowed	2.4 (3)	2.7 (3)
disallowed	0.0 (0)	1.8 (2)

[1]$R_{sym} = \Sigma|I_O-\langle I\rangle|/I_O$, where I_O is observed intensity, $\langle I\rangle$ is average intensity obtained from multiple observations of symmetry related reflections.

[2]The highest resolution shell was 2.7-2.8 Å for the apo BRC structure and 2.8-2.9 Å for the BRC-TPP complex.

[3]$R_{iso} = \Sigma||F_{PH}|-|F_P|| / \Sigma||F_P|+|F_{PH}||$, where $|F_P|$ is the protein structure factor amplitude, $|F_{PH}|$ is the heavy atom derivative structure factor amplitude.

[4]$R_{cullis} = \Sigma||F_{PH}\pm F_P|-F_{H(calc)}| / \Sigma|F_{PH}\pm F_P|$.

[5]Phasing power is $[\Sigma|F_{PH(calc)}|^2 / \Sigma\{|F_{PH(obs)}|-|F_{PH(calc)}|\}^2]^{1/2}$.

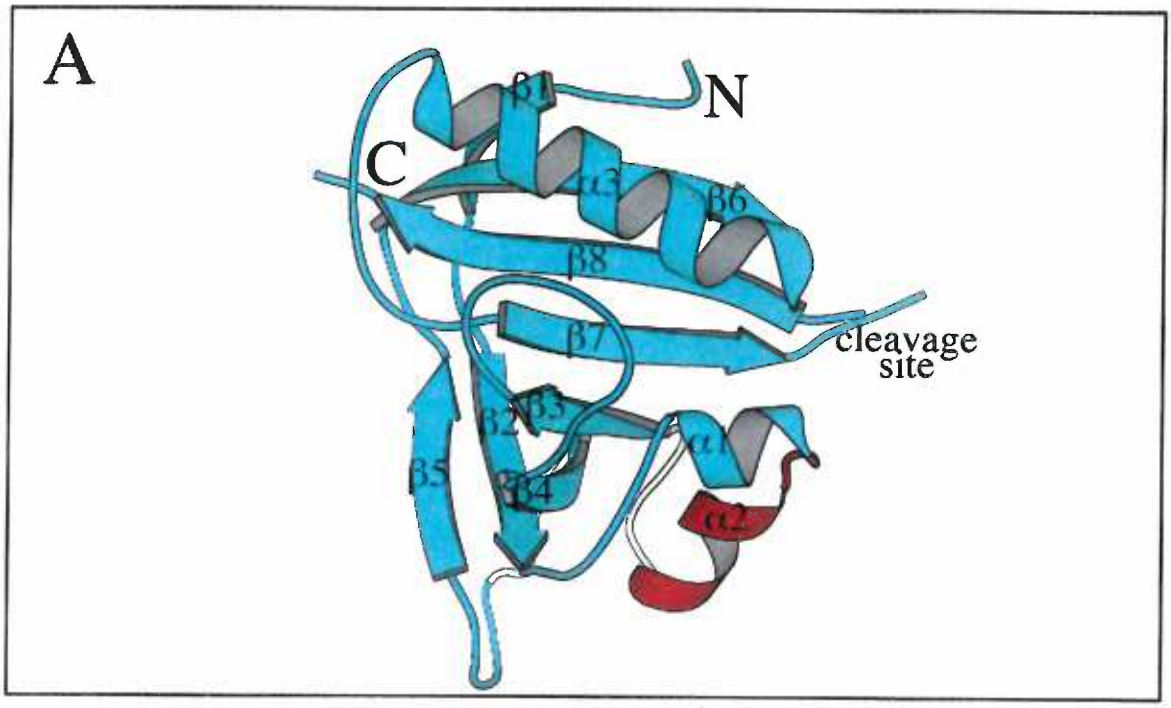
[6]Figure of merit is $\langle|\Sigma P(\alpha) e^{i\alpha}/\Sigma P(\alpha)|\rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution.

[7]$R_{factor} = \Sigma||F_{obs}|-|F_{calc}|| / \Sigma|F_{obs}|$.

Figure 4.1. Ribbon diagrams of the monomeric and dimeric apo BRC.

(A) Ribbon diagram of the BRC monomer with the secondary structure elements depicted as coils for α-helices, arrows for β strands and tubes for other structures. Labelled are the amino (N) and the carboxy (C) termini, and the trypsin cleavage site. The protein is composed of strand β1 (residues 6-10), strand β2 (residues 13-18), helix α1 (residues 24-30), helix α2 (residues 34-42), strand β3 (residues 51-55), strand β4 (residues 69-74), strand β5 (residues 89-93), strand β6 (residues 96-104), helix α3 (residues 107-123), strand β7 (residues 132-137), and strand β8 (residues 149-157). The disordered residues 142-147 are not shown. Loops that have poor electron density (residues 43-47 and 77-80) are shown in white. Helix α2 (residues 34-42), which becomes disordered upon TPP binding, is shown in red. This figure was generated with Molscript (Kraulis, 1991).

(B) Ribbon diagram of the BRC dimer. Labelled are the amino (N) and the carboxy (C) termini, and the drug-binding site residues Glu134 and Glu134'. Shown in black as ball-and-stick representations are the drug-binding site residue Glu134, the metal-chelating residues Glu19 and His70, and their two-fold symmetry mates. The metal ion coordinated at the dimer interface is shown as a yellow sphere. The metal ion has been refined as Mn^{2+} at occupancy 0.5 with a resulting thermal B-factor of 18 $Å^2$. This figure was generated with Molscript (Kraulis, 1991).


A
β1
N
C
α3
β6
β8
β7
cleavage
site
β2
β3
α1
β5
β4
α2

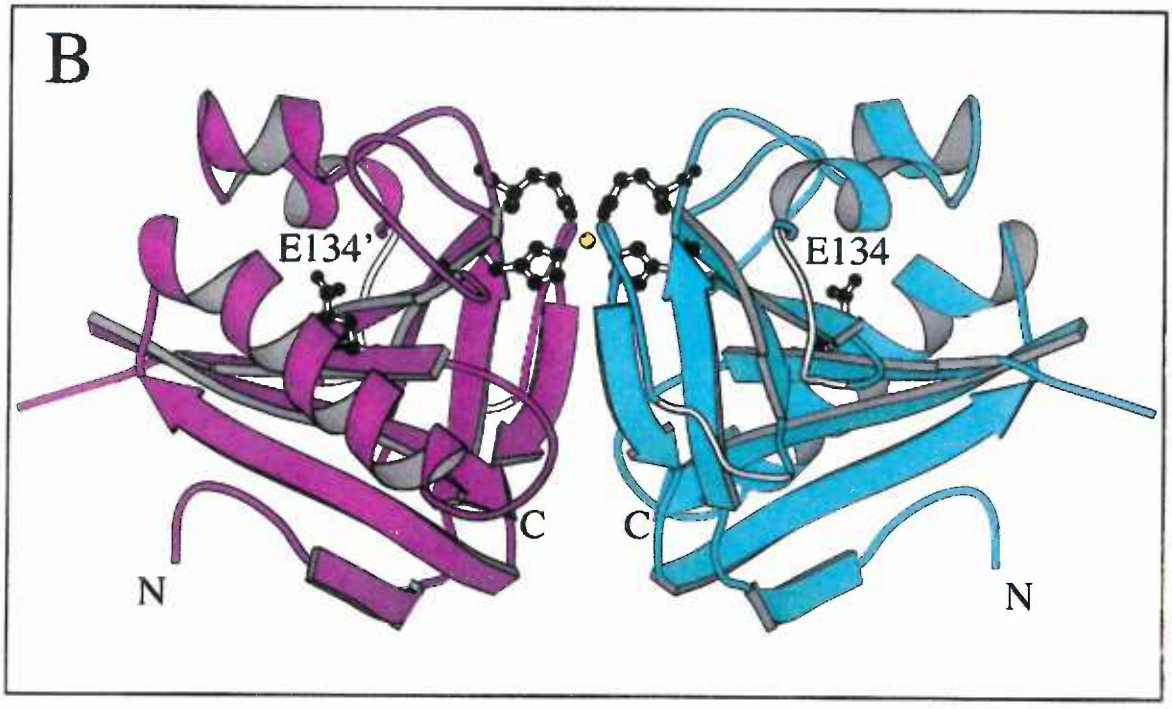

B
E134'
E134
C
C
N
N

Figure 4.2. An F_o-F_c "omit" map of apo BRC residues Glu134, Tyr33, Tyr68, and Tyr110 contoured at 3.5σ. Dashed lines indicate hydrogen bonds between the carboxylate moiety of Glu134 and hydroxyl groups of residues Tyr33, Tyr68, and Tyr110 with their distances given in Å. This figure was generated with O (Jones et al., 1991).

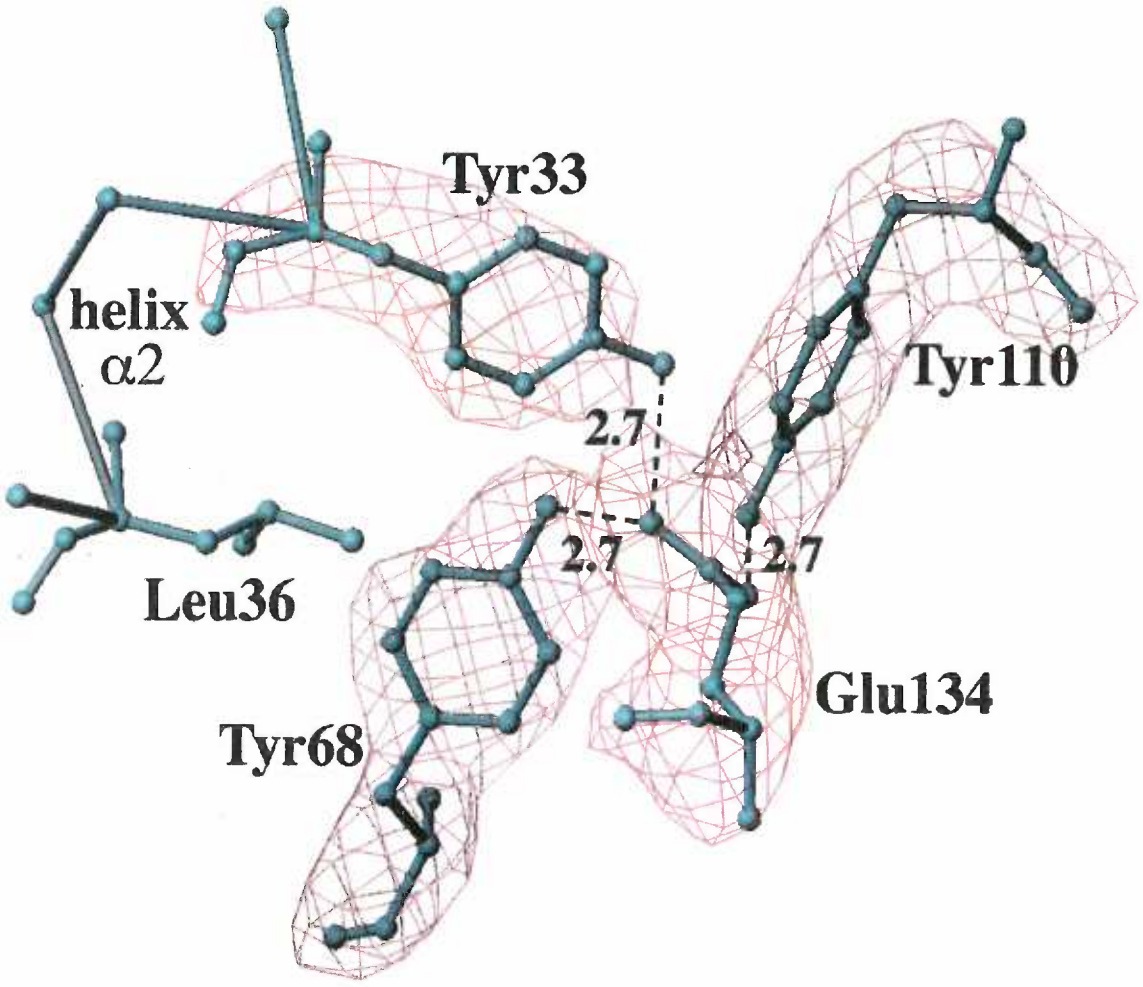

Tyr33
helix
α2
Tyr110
2.7
2.7
2.7
Leu36
Glu134
Tyr68

Figure 4.3. Structural changes in BRC upon TPP binding. Apo BRC dimer (top) and TPP-bound BRC (bottom) are represented as cyan ribbons. Upon drug binding, residues 28-50 (shown in red) undergo significant conformational changes. A dotted line depicts disordered residues 34-49. The metal-chelating residues Glu19, His70, Glu19′, and His70′ are shown as cyan balls and sticks, and binding-pocket residues Glu134 and Glu134' are depicted as blue balls and sticks. The bound TPP molecules are shown as purple balls and sticks and the metal ion is shown as a red sphere. This figure was generated with Sybyl (Tripos).

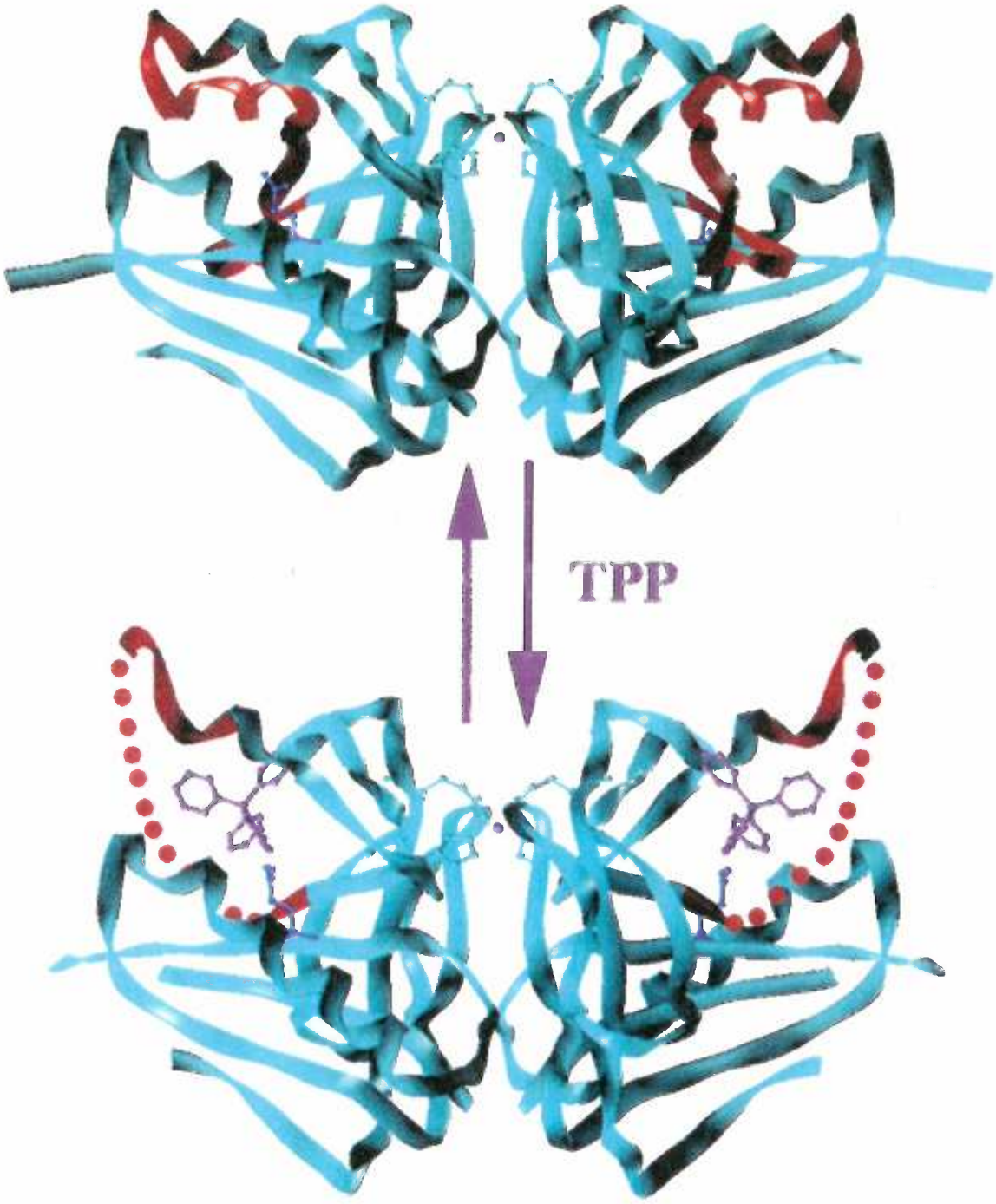

TPP

Figure 4.4. The F_o-F_c "omit" map of the tetraphenylphosphonium molecule contoured at 2.5σ. BRC residues making contact to the TPP and water molecule Wat1 are shown in cyan. The TPP molecule is shown in red. Dashed lines indicate hydrogen bonds and contacts between selected atoms with distances given in Å. This figure was generated with O (Jones et al., 1991).

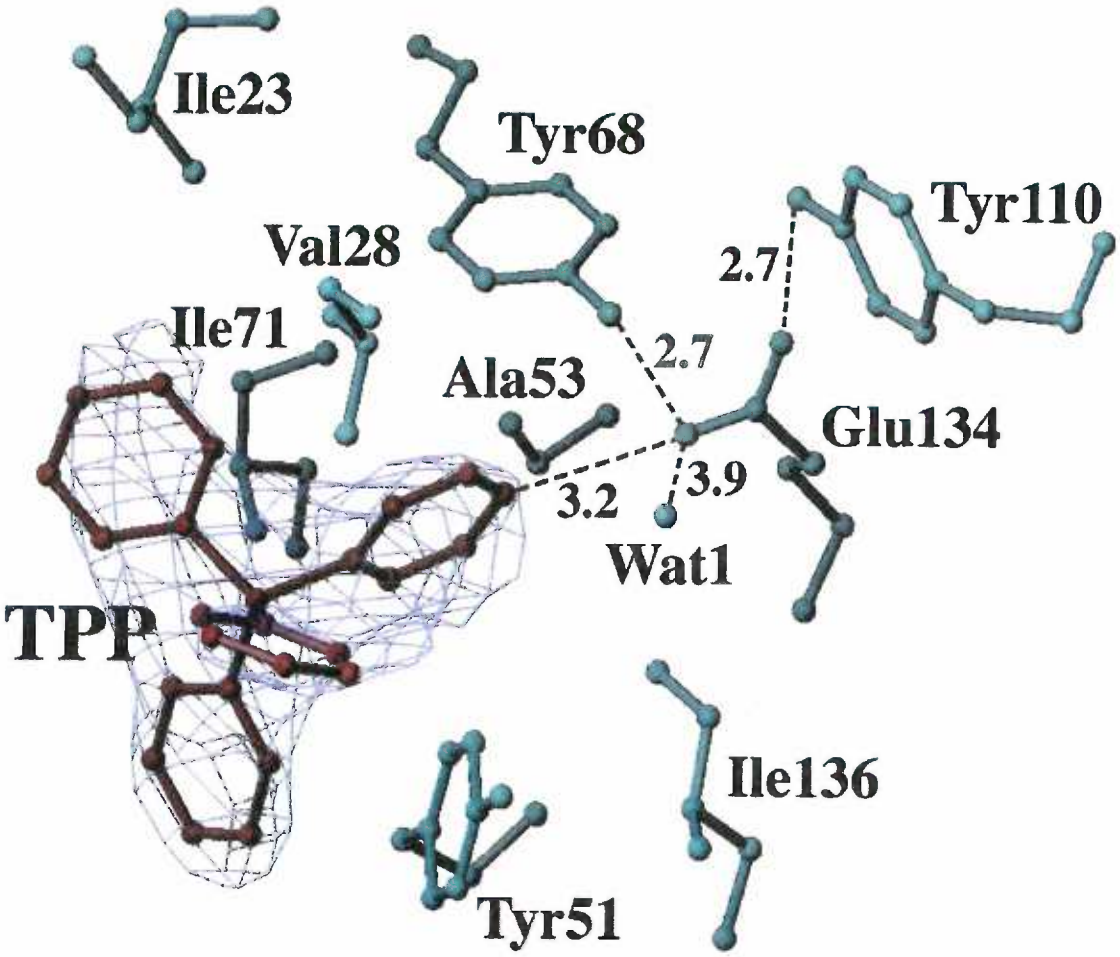

Ile23
Tyr68
Tyr110
Val28
2.7
Ile71
2.7
Ala53
Glu134
3.2
3.9
Wat1
TPP
Ile136
Tyr51

Figure 4.5. The electrostatic surface potential of the monomeric apo BRC (left) and TPP-bound BRC (right). Regions of the negative potential are coloured red, positive, blue, and neutral, white. In apo BRC (left), residues 29-42 are shown as sticks where carbons are white, nitrogens blue and oxygens red. These residues are excluded from the surface calculations. In the TPP-bound BRC (right), a molecule of TPP is shown as a stick representation where the phosphorus is yellow and carbons are white. The TPP molecule is excluded from the surface calculations. This figure was generated with GRASP (Honig and Nicholls, 1995).

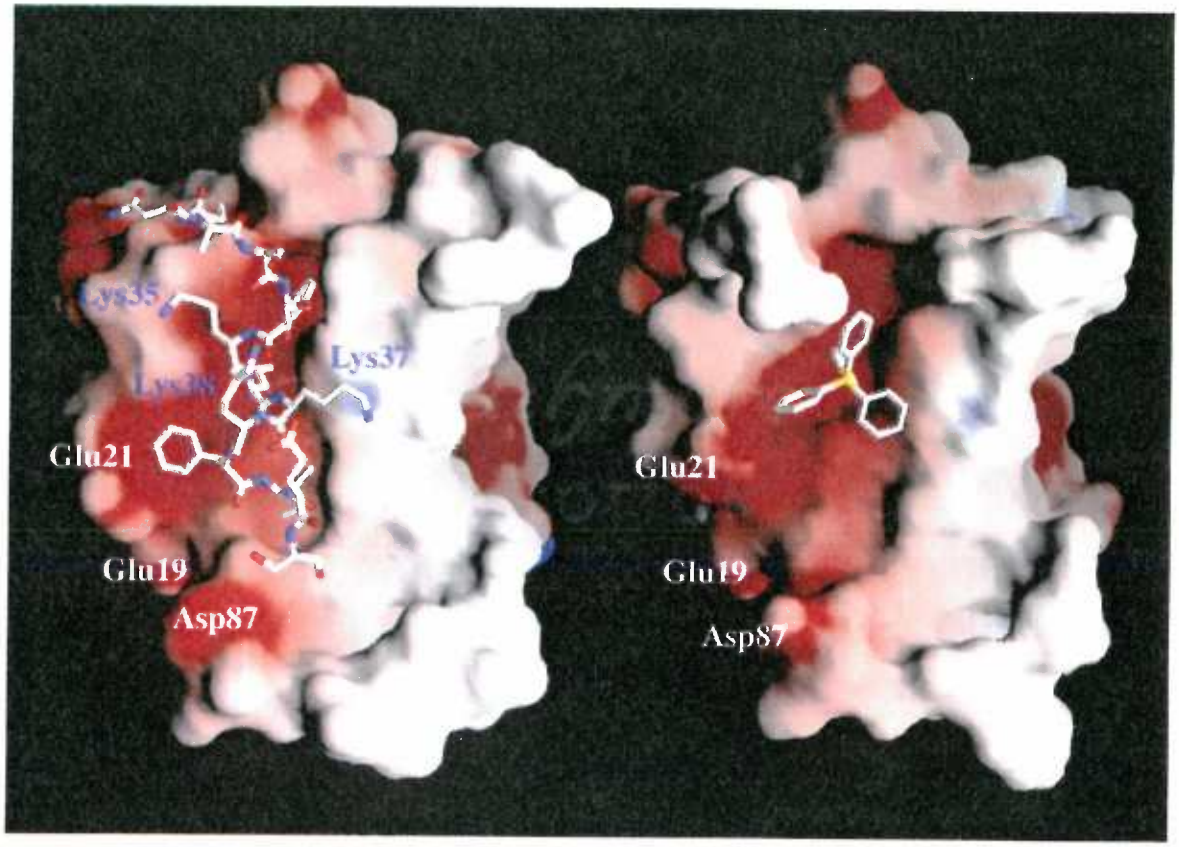
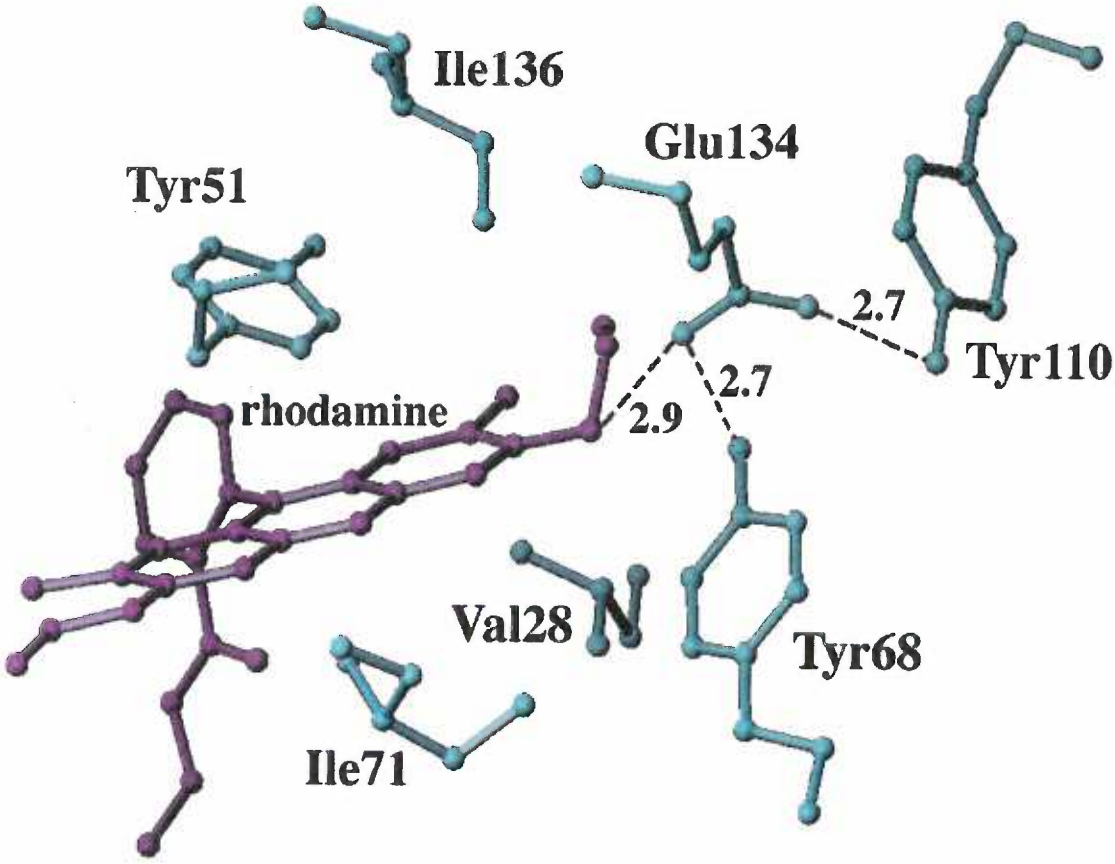

Lys35
Lys37
Glu21
Glu19
Asp87
Glu21
Glu19
Asp87

Figure 4.6. Structure-based model of rhodamine bound to BRC. Rhodamine is displayed as purple balls and sticks and the protein atoms as cyan balls and sticks. Dashed lines indicate potential interactions between selected atoms with distances given in Å. This figure was generated with O (Jones et al., 1991).


Ile136
Glu134
Tyr51
2.7
Tyr110
2.7
2.9
rhodamine
Val28 N
Tyr68
Ile71

CHAPTER 5

THE ROLES OF EXOCYCLIC GROUPS OF THE CENTRAL BASE-PAIR STEP IN MODULATING THE AFFINITY OF PURR FOR ITS OPERATOR: BIOCHEMICAL AND STRUCTURAL STUDIES.

Ekaterina E. Zheleznova and Richard G. Brennan§

Department of Biochemistry and Molecular Biology,
Oregon Health Sciences University, Portland, Oregon 97201-3098

§Author to whom correspondence should be addressed.
Phone: (503) 494-4427 FAX: (503) 494-8393
E-mail: brennanr@ohsu.edu

Abstract

E.coli purine repressor PurR kinks its cognate palindromic operators at the central CpG sequence or CpG step. PurR binding is sequence-specific, but remarkably, this conserved CpG step does not make any specific contacts to the protein. Yet, substitution of the TpA step for CpG abolishes PurR repression in vivo. To understand the requirements imposed on the central sequence by PurR, we challenged PurR with a number of systematically altered *purF* operators, in which the central CpG step (two dyad-related C·G base-pairs) was replaced with CpI, TpA, UpD, TpD, UpAP, TpAP, 5meCpG and UpiG. Specifically, we measured affinity of PurR for all modified *purF* operators and determined crystal structures of PurR in complex with UpD, 5meCpG, UpAP, and TpD. Our results pinpoint certain exocyclic groups in the kinked region as modulators of PurR/DNA interaction. In the major groove, the 5-methyl group on the pyrimidine interferes with the high-affinity binding because in the kinked *purF* operator it comes into an unfavourable steric apposition with the group at the exocyclic 6 position of a neighbouring purine thereby increasing the energetic cost of attaining a proper kinked DNA conformation in PurR·DNA complex. In the minor groove, the 2-amino group on the purine, present in the CpG operator, is absolutely required for binding, likely because it expands the minor groove and in this manner assists in folding of the hinge helix.

Introduction

DNA-binding proteins have varying specificities for their cognate sequences, depending on their functions. Stringent specificity is commonly attained through series of direct protein-DNA contacts, such as hydrogen bonds or hydrophobic interactions, and through changes in the DNA conformation. As a model system to study the specificity requirements, we used the *E.coli* purine repressor PurR, which kinks its cognate operator in the central region.

PurR is a member of LacI repressor family and a key regulator of *de novo* purine and pyrimidine nucleotide biosynthesis in *E.coli* [He et al., 1990; Choi and Zalkin, 1990; Wilson and Turnbough, 1990; Rolfes and Zalkin, 1990b; He et al., 1993]. In the presence of corepressors, hypoxanthine or guanine [Rolfes and Zalkin, 1990a; Choi and Zalkin, 1992; Schumacher et al., 1995], this dimeric protein represses transcription from multiple genes by binding to its cognate 16 bp pseudopalindromic operators, the highest affinity site being the *purF* operator [Rolfes and Zalkin, 1988].

Previously, we determined the three-dimensional crystal structure of a PurR bound to its *purF* operator and to hypoxanthine found that this protein binds to its pseudopalindromic operator as a dimer; the 60-residue N-terminal DNA-binding domain of each monomer contacts an operator half-site [Schumacher et al., 1994]. The first DNA binding element consists of residues 4-23 folded into a helix-turn-helix motif, which binds in the major groove of the DNA. The second DNA binding element of PurR consists of residues 48-56, which compose the hinge helix. In the PurR dimer, two symmetric hinge helices wedge into the minor groove at the center of the pseudopalindromic binding site with the following sequence (central step is shown in bold):

5'-A2 A3 G4 A5 A6 A7 A8 **C9 • G9'**T8'T7'T6'G5'C4'G3'T2'-3'

3'-T2'T3'C4'T5'T6'T7'T8'**G9'• C9** A8 A7 A6 C5 G4 C3 A2 -5'.

Because the hinge helices are too large to fit into the 5.7 Å-wide minor groove of B-DNA, they expand the minor groove and significantly distort the *purF* operator by intercalating the side chains of symmetry-related Leu54 residues into the central Cyt9pGua9'(CpG) base-pair step of the *purF* operator. Locally, two central C·G base-pairs are unstacked, undertwisted and rolled out to 48°, and overall, the operator is sharply kinked away from the protein.

The essential central CpG step is conserved among all PurR operators [He et al., 1990]; in fact, a natural substitution of TpA step for CpG step results in a complete loss of repression *in vivo*. Yet, in the crystal structure, PurR makes no specific contacts (hydrogen bonds or specific hydrophobic interactions) to the CpG sequence. Instead, hinge-helix residue Leu54 makes van der Waals contacts to the O2 atom of Cyt9 (3.4 Å), to the deoxyribose O4' atom of Gua9'(3.3 Å), to the N3 atom of Gua9'(3.9 Å), and to the N2 group of Gua9' (4 Å). Of the four contacts, only the last one cannot be made in the TpA step. However, alone this contact is unlikely to account for the high specificity of PurR for the central CpG step.

How then does PurR distinguish the CpG step of the *purF* operator from the TpA step? To answer this question, we compared several chemical differences between the CpG and the TpA steps and designed base-analogue substitutions of C and G. To estimate the contribution of specific groups and features, we designed a series of *purF* operators in which the central CpG base-pair step was modified to CpI, TpA, UpD, TpD, UpAP, TpAP, and 5meCpG. Next, we challenged PurR with modified *purF* operators in a series of binding experiments and measured the affinities. In addition, we solved crystal structures of several PurR-hypoxanthine-modified *purF* complexes because we anticipated loss or gain of binding affinity to be correlated with the global and local structural changes found in these complexes. Here we show that in an indirect way, the purine 2-amino group and pyrimidine 5-methyl group dramatically affect the ability of PurR to recognize the central step of its operator.

Results

Design of modified *purF* oligonucleotides

To better understand why PurR discriminates against the central TpA base-pair step in absolute favour of the CpG base-pair step, we initially compared the chemical composition of CpG and TpA steps and dissected the chemical differences between them with particular emphasis on exocyclic groups. As seen in Figure 5.1, three distinct chemical differences between the bases in the C·G and T·A bases exist. First, guanine, unlike adenine, has an amino group at the exocyclic 2 position, and thus, CpG step has two exocyclic 2-amino groups in the minor groove. Second, cytosine, unlike thymine, lacks a methyl group at the exocyclic 5 position, and therefore, the major groove of the CpG step does not have any methyl groups. Finally, the C·G base-pair is stabilized by three Watson-Crick hydrogen bonds while there are only two hydrogen bonds to stabilize the T·A base-pair.

To determine the role of these exocyclic groups and hydrogen bonds in PurR specificity, we designed a series of systematically altered *purF* operators, where CpI, TpA, UpD, TpD, UpAP, TpAP, 5meCpG, or UpiG steps replaced the central CpG sequence (see Figure 5.2 and Figure 5.3 (left) for structures of corresponding base-pairs). The first set of operators, including CpI, TpD, and UpiG operators, probes the central step in the minor groove by deleting the 2-amino group in the CpG step (CpI operator), adding a 2-amino group in the TpA step (TpD operator), or replacing the amino group with the hydroxy group at the 2 position (UpiG operator). The second set of operators, including UpD, 5meCpG, and TpAP operators, tests the effect of the methyl group at position 5 of pyrimidine in the major groove. Finally, UpAP operator was used to examine the importance of 2 Watson-Crick hydrogen bonds.

DNA-binding studies by fluorescent anisotropy

The affinities of PurR for modified *purF* operators (TpA, CpI, mCpG, TpD, UpD, TpAP, and UpAP) were measured using fluorescence anisotropy, and the K_ds are listed in Table 5.1. Typical binding curves are shown in Figure 5.4 and Figure 5.5. Initially, we confirmed that PurR binds the native *purF* operator with a K_d of 2.5 nM as has been determined previously [Glasfeld et al., 1996]. Then, we tested the binding of PurR for the TpA operator but were unable to calculate the Kd of this binding because the binding curve was linear up to 5 μM PurR concentration. Such poor binding *in vitro* correlates with loss of PurR repression for the TpA-containing operator in *in vivo* repression assays.

Next, we tested the effect of 2-amino group on binding. If the 2-amino group is essential for PurR recognition, then by deleting this group from the CpG operator we would obliterate high-affinity binding. Indeed, when the 2-amino group was deleted from guanine (guanine ⇒ hypoxanthine substitution), PurR could bind the new CpI operator only non-specifically. Similar to the TpA operator, the binding curve was linear up to 5 μM PurR and the K_d of this binding could not be determined. Thus, upon the removal of the 2-amino group from guanine, the binding affinity dropped over 3 orders of magnitude. Such an effect indicated that the 2-amino group of guanine was absolutely required for PurR binding and its removal made a former CpG step functionally behave like a TpA step.

In the next experiment, we added a 2-amino group to the TpA step (adenine ⇒ 2,6-diaminopurine) to test whether a simple adding of the purine 2-amino group could restore the native binding affinity. The affinity of PurR for the TpD operator was higher than for the TpA operator (Kd of 1.2 μM), but was still about 600 times lower than the affinity for the native CpG operator. Thus, addition of 2-amino group alone could not completely restore the native affinity and suggested that some other group in the TpA base-pair step hinders the binding.

Our first candidate was the 5-methyl group on thymine because it is present in the low-affinity TpA and TpD operators but is missing from the native high-affinity CpG operator. Therefore, we modified the TpA step by simultaneously adding the 2-amino group to adenine (adenine ⇒ 2,6-diaminopurine) and removing the 5-methyl group from thymine (thymine ⇒ uracil). The UpD operator should functionally be very similar to the native CpG operator because both have the purine 2-amino groups and lack pyrimidine 5-methyl groups. As we had hypothesized, PurR bound the UpD operator with Kd = 2 nM. Thus, two modifications in the TpA step brought the complete restoration of the wild-type PurR affinity for its operator.

To confirm that the 5-methyl group on pyrimidine impairs PurR binding, we added the 5-methyl group to the cytosines in the native CpG operator and measured PurR affinity for this 5meCpG operator. As we expected, the addition of the methyl group dramatically weakened PurR binding by 550-fold (Kd = 1.1 μM). Clearly, the 5-methyl group somehow interferes with binding, yet it makes no contacts with PurR. To gain insight into the its mode of action, we used crystal structure of PurR in complex with its native recognition site to build the 5-methyl group into the CpG step. If the 5meCpG operator assumed the exactly same conformation as the native *purF* operator, the distance between 5-methyl group of cytosine *Cyt9* and the exocyclic 6-oxo group of guanine Gua9' would be only 2.7 Å (Figure 5.6). This distance is much shorter than the optimal van der Waals distance between those atoms and to avoid this clash between the methyl group of Cyt9 and the Gua9' O6, the DNA must undergo an energetically expensive readjustment. Similarly, in the TpD step the 5-methyl group of thymine Thy9 would clash with the amino group at the exocyclic 6 position of 2,6-diaminopurine Dap9'. Interestingly, the $\Delta\Delta G_{monomer}$ between TpD and UpD operators is the same as the $\Delta\Delta G_{monomer}$ between 5meCpG and CpG operators: 1.9 kcal/mole. Thus, the energetic cost of adding a 5-methyl group to pyrimidine in the central step of *purF* operator does not depend on the major groove environment, that is, an amino or an oxo group at the exocyclic 6 position of purine.

Furthermore, the modelling suggested that in the absence of a group at the exocyclic 6 position of a purine, 5-methyl group should not hinder binding. We removed the 6-amino group from 2,6-diaminopurine in the TpD operator (2,6-diaminopurine ⇒ 2-aminopurine). To determine the effect of 5-methyl group within this new environment, we measured binding of PurR to TpAP-substituted *purF* operator. According to our expectations, PurR bound TpAP operator with nearly wild-type affinity (K_d = 12 nM).

In the final experiment, we asked whether the presence of 3 base-pairing hydrogen bonds of a C·G base-pair was required for high-affinity binding of PurR. Both T·A and C·I base-pairs have only 2 Watson-Crick hydrogen bonds. Incidentally, the TpAP operator has tested this requirement because the removal of 6-amino group eliminated its hydrogen bond with 4-oxo group of thymine. The 6-fold difference in affinity of PurR for the CpG and the TpAP operators suggested that the presence of an additional base-pairing hydrogen bond plays only a small role in high-affinity PurR binding of its operator. To confirm this conclusion, we designed a UpAP operator by removing the 6-amino group from the 2,6-diaminopurine in the UpD operator (2,6-diaminopurine ⇒ 2-aminopurine), which PurR binds with wild-type affinity. Thus, the U·AP base-pair has only 2 hydrogen bonds but retains purine 2-amino group and lacks pyrimidine 5-methyl group. We detected only a slight, 3-fold, decrease in the PurR affinity for the UpAP operator (Kd = 6 nM) in comparison with the UpD operator. Hence, the presence of 2 base-pairing hydrogen bonds is sufficient for the high-affinity binding, but disruption of two Watson-Crick hydrogen bonds in the native operator (one hydrogen bond per base-pair) costs about 0.3 kcal/mole of $\Delta\Delta G_{monomer}$.

In short, 2 rules have emerged from the DNA-binding studies: PurR discriminates against operators that contain 5-methyl group in the major groove (with an exception of the TpAP operator) and lack the 2-amino group in the minor groove. Interestingly, the second requirement is more stringent than the first one; PurR is still able to bind operators containing the 5-methyl group (TpD and 5meCpG operators) (Figure 5.5), but in the

absence of the 2-amino group, the binding becomes nonspecific (CpI and TpA operators). Thus, with almost wild-type affinity, PurR binds only CpG, UpD, UpAP, and TpAP operators (Figure 5.4).

Structural analysis of PurR in complex with modified operators

To determine the detailed architecture of the modified *purF* operators in complex with PurR and to gain insight into the structural basis for the observed thermodynamic effects of the substitutions, we crystallized PurR bound to four modified *purF* operators, in which the central CpG step was replaced with UpD, UpAP, 5meCpG, or TpD steps. Subsequently, the crystal structures of the UpD, UpAP, 5meCpG, and TpD-containing complexes were solved to the resolution of 2.9 Å, 2.9 Å, 2.7 Å, and 2.7 Å, respectively. Crystallographic statistics is given in Table 5.2. Figure 5.7 shows the global architecture of PurR bound to TpD operator and hypoxanthine. Overall, the 4 structures are very similar to the native PurR-hypoxanthine-*purF*-operator complex structure with only minor differences that are discussed below. Superimposition of all protein Cα atoms and all DNA N1 atoms of each modified complex, containing UpD, UpAP, 5meCpG, or TpD, onto those of the native complex resulted in rmsds of 0.22 Å, 0.18 Å, 0.19 Å, and 0.26 Å, respectively.

DNA conformation

The DNA conformational parameters of the native and four modified *purF* operators, as determined by program CURVES, are listed in Table 5.3. Globally, the four structures are essentially identical (Figure 5.8), and small differences between the native and the modified DNA are within the uncertainty associated with the atomic positions of the models. As found in the native PurR-hypoxanthine-*purF* complex, the largest DNA deformations occur in the central CpG (Cyt9pGua9') step. To accommodate the

intercalating Leu54 side chain, the minor groove is widened (Figure 5.8), reducing the helical twist locally, and bases are unstacked, which results in the increased rise. As seen in Table 5.3, the local parameters for the C·G base-pair and for the CpG step are very similar in all these structures. However, there are several differences caused by the base modifications.

When the DNA conformation of the modified *purF* operators is compared to that of the native *purF* operator, the most prominent differences are found between the UpD operator and the native operator. This differences can be explained by examining the major groove. First, the roll of the UpD step (33.9°) is smaller than that of the CpG step (48°). In the native structure, the rolling out of the central step brings O6 of Gua9' within the 3.3 Å distance from the N6 of Ade8 (Figure 5.9A), which is favourable due to the formation of a hydrogen bond. The UpD operator, however, has an N6 group from 2,6-diaminopurine. Because of the steric clash, the 3.3 Å distance between Dap9' N6 and Ade8' N6 is not favourable, and the bases undergo several readjustments. A smaller roll angle decreases the rise between U·D and D·U base-pairs. The propeller twist of the neighbouring A·T base-pair changes in the positive direction from -19.1 to -14.8 slightly straightening the basepair; likewise, the propeller twist of the central base-pair changes from -7.9 of the C·G base-pair to 2.1 of the U·D base-pair. As a result, the N6 (Ade8) - N6 (Gua9') distance increases to a more favourable van der Waals distance of 3.7 Å (Figure 5.9B) [Li et al., 1998]. Despite the smaller roll angle in the UpD operator, the overall DNA bend (53.9°) is not affected as compared to the bend of the native operator (52.1°). In general, the roll angle does not correlate with the overall DNA bend. For example, SRY creates a 65° bend in its cognate DNA with only a 21° roll in the ApA step [Werner et al., 1996; Olson et al., 1998].

In the 5meCpG structure, the CpG step also undergoes conformational change to avoid the modelled clash between the meCyt9 C7 and Gua9' O6. Indeed, in the crystal structure, the distance between the C7 atom of meCyt9 and the O6 atom of Gua9' is 3.3 Å (Figure 5.9C). Modelling experiments have suggested that the DNA should either be less

bent or undergo other readjustments. In reality, the DNA chooses the second option possibly because the 52° bend is important for the formation and global stability of the PurR·DNA complex. To increase the distance between the otherwise clashing C7 of Cyt9 C7 and O6 of Gua9', the roll angle in the central CpG step decreases from 48° to 43.6° and the propeller twist of the 5meC·G base-pair becomes positive, which flattens the 5meC·G base-pair. Unlike in the UpD operator, in the 5meCpG operator, it is not necessary to flatten the A8·T8' base-pair, and the propeller twist of the latter base-pair (-19.6°) remains the same as that in the native operator (-19.1°).

In the TpD structure, the central step undergoes conformational changes similar to those in both the 5meCpG and the UpD structures. To avoid the clash between the C7 atom of Thy9 and the N6 atom of Dap9' (meCpG-like) and between the N6 atom of Ade8 and the N6 atom of Dap9' (UpD-like), the DNA undergoes several adjustments (Figure 5.9D). The roll angles in the central T·D/D·T step (TpD) decrease from 48.0° to 39.8° while the roll of the preceding A·T/T·D base-pair step decreases from 3.0° to -2.5°. The propeller twist of the T·D base-pair decreases from -7.9 to -1.1, which again flattens this base-pair. Likewise, the propeller twist of the A8·T8' base-pair, directly 5' of the central base-pair step, becomes more positive decreasing from -19.1 to -15.2 and thereby, slightly straightening the A8·T8' base-pair. As the result of all these rearrangements, the distances between the C7 (Thy9) and the N6 (Dap9') as well as between the N6 (Ade8') and the N6 (Dap9') increase to 3.6 Å (Figure 5.9D).

Despite the changes in the central base-pair step, the global conformations of the 5meCpG, UpD, and TpD-containing operators are very similar to that of the native *purF* operator (Figure 5.8), indicating that locally, the DNA is plastic enough to adopt the conformation appropriate for PurR binding. This can be illustrated by similar values for the global bend, which are 52.1, 53.9, 53.8, and 51.6 for the CpG, UpD, 5meCpG, and TpD operators, respectively (Table 3). In all operators, the remainder of the operator

quickly recovers the B-DNA geometry returning to appropriate helical twist, rise, and minor groove width within two base steps of the Leu intercalation site.

The structure of the UpAP-containing complex is very similar to the structure of the native *purF* operator (Figure 5.9E), which indicates that the presence of 3 base-pairing hydrogen bonds plays only a small role in determining the conformation of the DNA.

PurR·operator contacts in the minor groove

In the native PurR-hypoxanthine-*purF* operator structure, specificity is provided by residue Lys55, which makes a weak hydrogen bond from its ε-NH_3 to the N3 of Ade8 and van der Waals contacts to C2 (3.2 Å). Similar contacts have been found in all modified complexes. Hinge-helix residue Leu54 makes van der Waals contacts to the O2 atom of Cyt9 (3.4 Å), to the deoxyribose O4' atom of Gua9'(3.3 Å), to the N3 atom of Gua9'(3.9 Å), and to the N2 group of Gua9' (4.0 Å). The examination of modified complexes revealed similar contacts between PurR and the modified DNAs.

Discussion

The role of 5-methyl group

The presence of 5-methyl group on pyrimidine of the central step significantly impairs the DNA-binding ability of PurR. We found that PurR binds with a 600-fold higher affinity to the operator in which the central step lacks the 5-methyl group. This directly implies that PurR prefers a CpG step over a 5meCpG step and a UpD step to a TpD step. Prior to all crystal structure determinations, modelling studies suggested a likely reason for the observed changes in the binding affinity. Namely, if the 5meCpG and TpD operators maintain the same conformation as the native PurR-bound *purF* operator, the 5-

methyl group of the pyrimidine (5meCyt9 or Thy9) would come into an unfavourable proximity (2.7 Å) with either the O6 group on Gua9' or the N6 group on Dap9' (Figure 5.6). In both complexes, the central operator step undergoes a similar conformational change, which relieves such a steric strain and results in the 3.3 Å distance between the C7 atom of meCyt9 and the O6 atom of Gua9' (in 5meCpG operator) (Figure 5.9C) and the 3.6 Å distance between the C7 atom of Thy9 and the N6 atom of Dap9' (TpD operator) (Figure 5.9D). Both distances are well within accepted van der Waals contact distances [Li et al., 1998].

Clearly, this steric strain does not occur in the central step that has a pyrimidine 5-methyl group but lacks an exocyclic (either oxo or amino) group at the exocyclic 6 position in the purine ring, such as TpAP operator. Indeed, we found the affinity of PurR for the TpAP operator to be very close to that for the CpG operator, which supports our conclusion that a 5-methyl group diminishes the binding only when it can create a steric clash in the major groove. The small 6-fold decrease in the binding affinity can be explained almost completely by the loss of a hydrogen bond between the 6-oxo group of guanine and the 4-amino group of cytosine.

Thus, the pyrimidine 5-methyl groups in the 5meCpG, TpD, and TpA step impairs binding because DNA-bending by PurR would result in a steric strain in the major groove, which can be alleviated only by thermodynamically costly conformational changes in the central step. From our binding data, the energetic cost of adding an exocyclic methyl group to the pyrimidine is 1.9 kcal/mole. Despite a large drop in affinity, the addition of a 5-methyl group does not affect the global structure of the PurR·DNA operator complex, as indicated by nearly identical global DNA bend angles (Tables 3).

The role of 2-amino group

Our binding experiments have demonstrated that the amino group at the exocyclic 2 position of the purine is required for the high-affinity binding of PurR; its removal obliterated PurR binding. Why is 2-amino group necessary for PurR binding?

Due to the lack of structural information, we cannot directly compare the DNA conformation and the PurR-DNA contacts in complexes that have the 2-amino group with those that lack it. Thus, we attempted to estimate the contribution of 2-amino group by analyzing the contacts it makes in crystal structures. First, this 2-amino group makes a hydrogen bond to 2-oxo group of the pyrimidine in C·G and U·D base-pairs, which is lost when 2-amino group is removed. However, such loss can only account for a 0.3 kcal/mole change in $\Delta\Delta G$, calculated for the loss of a comparable hydrogen bond between the purine 6-amino group and the pyrimidine 4-oxo group.

Second, in the native structure, the 2-amino group makes a single van-der-Waals contact (4.0 Å) to the Cδ2 atom of the intercalating residue Leu54 of PurR. The same contact has been found in all four crystallized complexes, varying from 3.5 Å to 3.9 Å. When the 2-amino group is removed, this favourable contact is lost, and in the core of the PurR-DNA complex, a small cavity is created. Such a cavity is energetically unfavourable and should decrease the stability of PurR-DNA complex. Indeed, in thermodynamic studies on the I41V mutant of Arc protein, the δ-CH_3 groups of Ile were shown to contribute 1.5-2.0 kcal/(mole of dimer) to the protein stability [Milla et al., 1995]. Also, a study on Ile->Val mutations in chymotrypsin estimated the average change in the free energy of unfolding for deleting a methylene group as 1.3+/- 0.5 kcal/mole [Jackson et al., 1993]. Yet, this value is too small to explain the loss of PurR affinity for operator sites without a 2-amino group.

The key question is whether in the absence of the 2-amino group PurR is able to form its specific complex with the operator. Our inability to crystallize PurR in complex

with either TpA or CpI-containing *purF* operators suggests that upon the deletion of the purine 2-amino group PurR·*purF* complex cannot form. Interestingly, low affinity alone does not prevent crystallization, as the PurR·TpD and PurR·5meCpG complexes (the Kds of 1.2 and 1.1 μM, respectively) crystallize under the native conditions. Moreover, the deletion of the minor groove side chain does not prevent a crystallization either. A Lys55Ala substitution creates a cavity in the core of the PurR·DNA complex and decreases the Kd to 0.9 μM, yet the mutant crystallizes under the same conditions and in the same conformation as the wild-type protein [Glasfeld et al., 1996]. In addition, several cavity-generating mutants of lysozyme have been previously crystallized, and their structures have been solved [Eriksson et al., 1992].

To test whether PurR forms a specific complex with an operator, one can digest protein in the presence of the operator. It is known that the hinge helix residues are disordered in the corepressor-free conformation [Choi and Zalkin, 1994]; however, corepressor binding brings the disordered hinge regions of each monomer together and allows them to fold into helices in the presence of DNA. Proteolysis studies [Choi and Zalkin, 1994] have shown that the hinge helix region is highly susceptible to cleavage between residues Arg52 and Ser53 only when PurR in not specifically bound to DNA – a strong indication that DNA is mandatory for the stable formation of hinge helices. Furthermore, NMR studies on the DNA-binding domain of PurR have demonstrated that in the absence of DNA, hinge residues form no discernable secondary structure [Nagadoi et al., 1995].

Thus, to determine whether PurR forms a specific complex in the absence of the 2-amino group, we digested PurR with trypsin in the presence of hypoxanthine and CpI operators, using 5meCpG and CpG operators as controls. We hypothesized that if PurR bound the CpI operator specifically, the hinge helix would be folded and trypsin would not be able to cleave off the DNA-binding domain. Our results were convincing: PurR remained intact when incubated in the presence of the CpG or 5meCpG operators but was

broken down into the DNA-binding and the corepressor-binding fragments in the presence of the CpI operator. Thus, because the operator serves as a template for the local folding in the protein, the 2-amino group appears to be necessary for the folding of hinge helix or at least its stabilization that is disordered in apo PurR.

Next, we asked how the 2-amino group could possibly affect the formation and stabilization of hinge helices. A possible explanation came from a variety of studies on the role of 2-amino group in DNA conformation [Bailly et al., 1995]. In particular, the 2-amino group of guanine widens the minor groove of G·C base-pairs as compared with that of A·T base-pairs [Dickerson et al., 1994], where the exocyclic 2-amino group on a purine is not present. Thus, PurR is likely to need a wider minor groove to assist the folding of the hinge-helix residues and their precise fit into the minor groove of an operator.

At present, it is unclear whether any group at the exocyclic position 2 of a purine is sufficient for the formation of hinge helices or whether PurR selects the 2-amino group at this position. So far, we have designed a UpiG operator (uracil·isoguanine), which has a hydroxy group at the exocyclic position 2 of the purine. However, the UpiG base-pair have been shown to exist in two tautomeric forms (Figure 5.3) [Robinson et al., 1998]. Whereas one form has Watson-Crick base-pairing, reminiscent of the CpG base-pairing, the other form exists in a "wobble" conformation, unsuitable for PurR binding. Unfortunately, the "wobble" tautomer is the dominating species under conditions of our DNA-binding experiments, as has been confirmed by clearly different melting temperatures for the *purF* and the UpiG operator DNA, 54°C and 47°C , respectively (data not shown).

Conclusions

Our results suggest that the geometry of PurR-operator complex imposes a set of specific requirements for exocyclic groups in the major and minor grooves of the central YpR step. Having analyzed crystal structures of PurR complexed to 4 modified operators and a native operator, we conclude that to form the high-affinity complex, PurR and its

cognate DNA must be able to adapt a specific global structure, in which the DNA is kinked by ~50° away from the protein and towards the major groove. The specific contacts along the entire protein-DNA interface (hinge helix-minor groove and recognition helix-major groove) necessitate such kinked conformation of the DNA. The affinity of PurR depends on the ability of DNA to assume this conformation and on the ability of the protein to it depends on protein-DNA interactions, responsible for a very snug fit of protein into its operator.

First, in the minor groove, the 2-amino group on the purine base is indispensable for specific binding. We hypothesize that the 2-amino group facilitates the formation of PurR/DNA complex by widening the minor groove and providing a template for the coil-to-helix transition of the hinge-helix residues. In its absence, the hinge helices do not fold, and hence, the specific PurR-operator complex cannot form. However, it is unclear whether the chemical nature of the group in the exocyclic 2 position is important. Second, PurR contacts its operator in the minor groove and bends it towards the major groove. Such compression of the major groove will bring into proximity any group at the exocyclic 5 position of the pyrimidine and at the exocyclic 6 position of a purine from the neighbouring base-pairs. The resulting steric strain imposes a limitation on the identity of the central base-pairs. We find that for high-affinity binding, one of these two positions must be vacant. The bulk of the methyl group of the 5meCpG and TpD operators provides a steric barrier to the compression of the major groove that occurs when PurR kinks the DNA. Thus, additional energy is required to overcome this energy barrier. All this results in the complete discrimination against a central TpA step for PurR as well as all other LacI/GalR family members.

Materials and Methods

PurR purification

PurR was expressed in *E.coli* and purified as described [Rolfes and Zalkin, 1990a].

Fluorescent anisotropy

The affinity of PurR for its native and modified *purF* operators in the presence of hypoxanthine was measured by fluorescence anisotropy [Lundblad et al., 1996] using a Beacon Fluorescence Polarization instrument (Panvera). Each DNA duplex was prepared from 2 complementary 24-mers, one of which contained a 5'-fluorescein label (Genosys and Oligos etc.). The sequence of the labelled *purF* operator nucleotide is 5'-F-GAAAAAGAAAAYRTTTGCTTAGGG-3' where the binding site is underlined. The central YpR step contained an appropriate modification of the CpG step and was TpA, CpI, 5meCpG, TpD, UpD, TpAP, or UpAP. The desired modifications were obtained by substitution of the appropriate modified nucleoside phosphoramidite during chemical oligodeoxynucleotide synthesis. Annealing of the complementary 5'-fluoresceinated and complementary non-fluoresceinated strands was accomplished by heating at 90° C for 5 minutes followed by slow cooling at room temperature. As judged by native polyacrylamide gel electrophoresis, single stranded oligonucleotides were absent from the preparation. The fluorescent anisotropy measurements were carried out at ambient temperature in 1mL of binding buffer (100 mM HEPES, potassium salt, pH 7.5; 250 mM potassium glutamate; 150 mM sodium chloride; 10 mM magnesium acetate; 0.5 mM EDTA, sodium salt; and 5% glycerol), which also contained 1μg poly dI:dC, 100mM hypoxanthine (a saturating concentration), and 2nM of labelled DNA duplex. PurR was titrated into the mixture, and measurements were done after a 30-second equilibration. The binding data were analyzed and fitted by nonlinear least-squares regression (Sigma Plot) using the equation:

$$\Theta = \frac{A-A_0}{Amax-A_0} = \frac{[P]}{Kd + [P]},$$

Where Θ is the fractional saturation, A is anisotropy, Amax is maximum anisotropy, A_0 is initial anisotropy, Kd is the dissociation constant, and [P] is the total protein concentration.

EDTA, sodium salt; and 5% glycerol), which also contained 1μg poly dI:dC, 100mM hypoxanthine (a saturating concentration), and 2nM of labelled DNA duplex. PurR was titrated into the mixture, and measurements were done after a 30-second equilibration. The binding data were analyzed and fitted by nonlinear least-squares regression (Sigma Plot) using the equation:

$$\Theta = \frac{A-A_0}{Amax-A_0} = \frac{[P]}{Kd + [P]},$$

Where Θ is the fractional saturation, A is anisotropy, Amax is maximum anisotropy, A_0 is initial anisotropy, Kd is the dissociation constant, and [P] is the total protein concentration.

Crystallization

In crystallization experiments, we used a 16 base-pair *purF* operator site with a 5'overhanging nucleoside. The sequence of one of the complementary *purF* operator strands is 5'-A1A2A3G4A5A6A7A8Y9•R9'T8'T7'T6'G5'C4'G3T2'-3', where YpR step was UpD, UpAP, TpD, or 5meCpG. We set up crystallization trials with all seven oligonucleotides. However, we successfully crystallized only four complexes of PurR with modified *purF* operators, namely, those in which the central CpG step has been replaced with UpD, UpAP, 5meCpG, or TpD. The crystals of PurR-hypoxanthine-modified *purF* complexes grew under the same conditions and are isomorphous with crystals of the PurR-hypoxanthine-native *purF* [Schumacher et al., 1994].

X-ray data collection and processing

X-ray intensity data for complexes containing UpD, UpAP, or mCpG central base step were collected to 2.9 Å, 2.9 Å, and 2.7 Å resolution, respectively, at room temperature with an Area Detector Systems Corporation (ADSC) multiwire area detector (Xuong et al., 1985) using a Rigaku RU200-H rotating anode X-ray generator with graphite

monochromator (40 kV, 150 mA). The data were processed with software provided by ADSC. X-ray intensity data for the complex containing TpD were collected to 2.7 Å resolution at room temperature on an R-AXIS IV imaging plate system using a RIGAKU RU300 rotating anode X-ray generator equipped with double focusing mirrors and operating at 50 kV and 100 mA. The data were processed using BIOTEX (Molecular Structure Corporation, Inc., Woodlands, TX).

Structure solution and model refinement

Three structures of PurR repressor bound to different modified *purF* operators were determined by difference Fourier methods because the crystals were isomorphous to the native crystals. The coordinates of the native PurR-hypoxanthine-*purF* complex were used as the starting model, but the Cyt9pGua9' step was replaced with Ura9pDap9', Thy9pDap9', Ura9pApu9', or Mcy9pGua9' to account for corresponding modifications.

The initial models underwent rigid-body refinement as implemented in software package TNT [Tronrud et al., 1987; Tronrud, 1997] with data from 10.0 Å to 3.5 Å. After convergence, the resolution was extended to 3.0 Å and positional least-squares refinement was carried out. After model rebuilding in O [Jones et al., 1991], the resolution was extended to 2.9 Å for UpD complex, to 2.9 Å for UpAP complex, to 2.7 Å for 5meCpG complex, and to 2.7 Å for TpD complex and positional and tightly restrained thermal-parameter refinement was carried out until the convergence. Solvent molecules were positioned into the overlapping 2Fo-Fc and Fo-Fc densities found within 3.4 Å distance from a hydrogen donor or acceptor. Locations of waters were verified by their removal followed by refinement until convergence and inspection of resulting Fo-Fc omit electron density maps. Only solvent molecules with B-factors less than 80 were included in the final model. The final models were verified by a series of Fo-Fc omit electron density maps in which 10% of the model was omitted and the remaining structure refined until convergence.

The stereochemistry of all four models is very good as shown by program PROCHECK [Laskowski et al., 1993]. There are only two Ramachandran plot outliers, Ser124 and Asp 275, in every model. The same residues were found to be outliers in the native PurR-hypoxanthine-*purF* structure. Relevant crystallographic statistics are presented in Table 2. Coordinates for the TpD complex (*1zay)* have been deposited to the Protein DataBase. Coordinates for the rest of the complexes have yet to be deposited.

Table 5.2. Crystallographic data.

	UpD	UpAP	meCpG	TpD
Resolution (Å)	27.3-2.9	42.5-2.9	26.3-2.7	60-2.7
Total observations	53,034	59,790	60,342	51,340
Unique reflections	14,304	15,499	18,425	17,098
R_{sym} (%)[1]	7.4	7.9	6.9	7.1
Completeness, %	93	100	96	88
Overall I/σ(I)	6.4	7.0	7.3	14.2
Highest resolution shell[2]				
Completeness, %	62	100	92	75
I/σI	1.4	1.7	1.5	2.8
Refinement statistics				
Resolution range (Å)	10.0-2.9	10.0–2.9	10.0-2.7	10.0-2.7
Reflections used in refinement	14,082	15,080	18,036	16,717
Completeness, %	93	100	96	88
Amino acid residues	338	338	338	338
Solvent molecules	33	41	31	25
R_{factor}[3] (%)	15.7	16.1	17.8	18.0
RMS deviations				
Bond length (Å)	0.012	0.011	0.012	0.010
Bond angles (°)	1.289	1.136	1.321	1.144
B-factor (Å^2)	5.82	4.72	5.36	4.64
PROCHECK: Residues in Ramachandran regions (%, number of residues)				
most favourable	250 (84.2)	259 (87.2)	259 (87.2)	256 (86.2)
allowed	43 (14.5)	34 (11.4)	34 (11.4)	38 (12.8)
generously allowed	2 (0.7)	2 (0.7)	2 (0.7)	1 (0.3)
disallowed	2 (0.7)	2 (0.7)	2 (0.7)	2 (0.7)

Table 5.2. Footnotes.

[1]$R_{sym} = \Sigma|I_O - \langle I \rangle| / I_O$, where I_O is observed intensity, $\langle I \rangle$ is average intensity obtained from multiple observations of symmetry related reflections.

[2]The highest resolution shell was 2.7-2.8 Å for the 5meCpG and TpD complexes and 2.9-3.0 Å for the UpD and UpAP complexes.

[3]$R_{factor} = \Sigma||F_{obs}| - |F_{calc}|| / \Sigma|F_{obs}|$.

Table 5.3. Summary of the DNA helical parameters (CURVES) CpG/UpD/meCpG/TpD/UpAP

Global inter base pair parameters	706 AT 707 CG	707 CG 708 GC
helical twist	14.6/20.8/17.5/25.5/21.3	22.1/28.4/29.3/24.7/30.9
Roll	3.0/7.1/7.8/-2.5/6.7	48.0/33.9/43.6/39.8/40.6
Rise	2.4/3.0/2.9/3.5/2.6	6.5/5.4/5.8/5.3/5.9
Tilt	11.4/2.6/6.0/4.8/4.1	0/0/0/0/0
Intra base pair parameters	**706 AT**	**707 CG**
X displacement	0.3/-1/-0.6/-0.9/-0.8	-0.2/-1.6/-1.1/-1.3/-1.2
Propeller twist	-19.1/-14.8/-19.6/-15.2/-17.5	-7.9/2.1/2.5/-1.1/-2.1
Buckle	12.1/14.7/14.1/16.6/15.0	23.1/11.7/14.6/13.1/17.6
Global bend		52.1/53.9/53.8/51.5/51.6

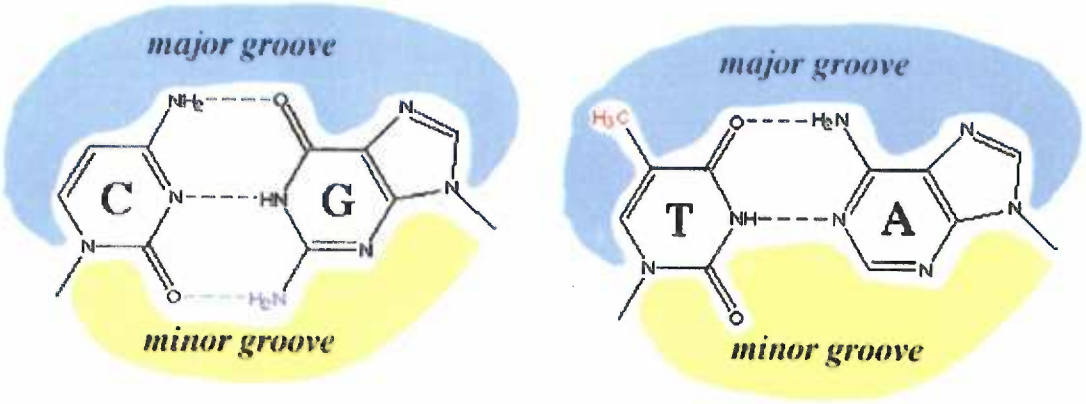


Figure 5.1. C·G (left) and T·A (right) base-pairs. Major grooves and minor grooves are labelled and coloured blue and green, respectively. The methyl group at the exocyclic 5 position of thymine is shown in red. The amino group at the exocyclic 2 position of guanine is shown in purple. The third Watson-Crick hydrogen bond of the C·G base-pair, which is absent in a T·A base-pair, is shown as a green dash line.

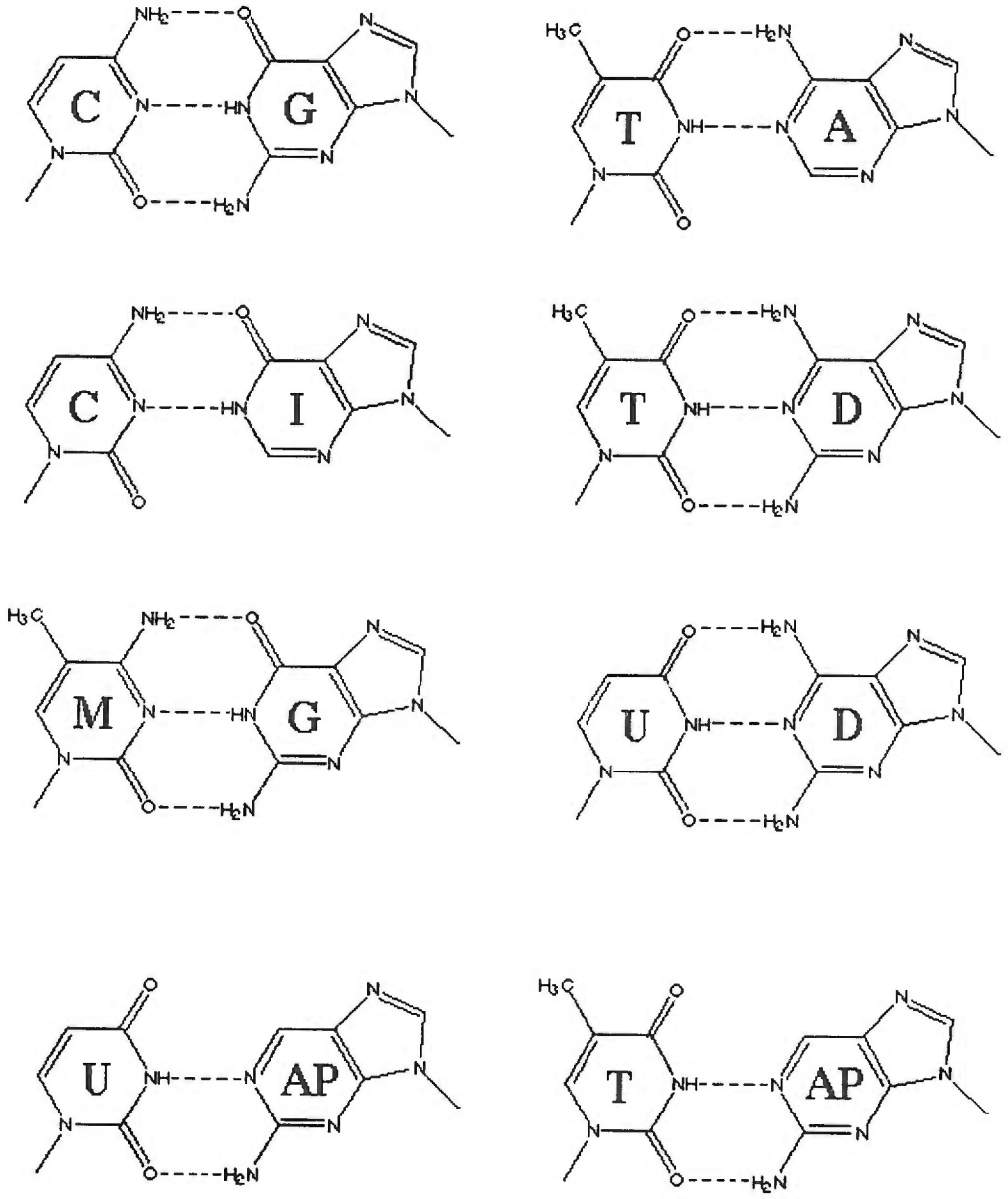


Figure 5.2. Structures of hydrogen-bonded pyrimidine-purine base-pairs used in this work. D represents 2,6-diaminopurine, AP represents 2-aminopurine, iG represents isoguanine, I represents inosine, and M represents 5-methylcytosine.

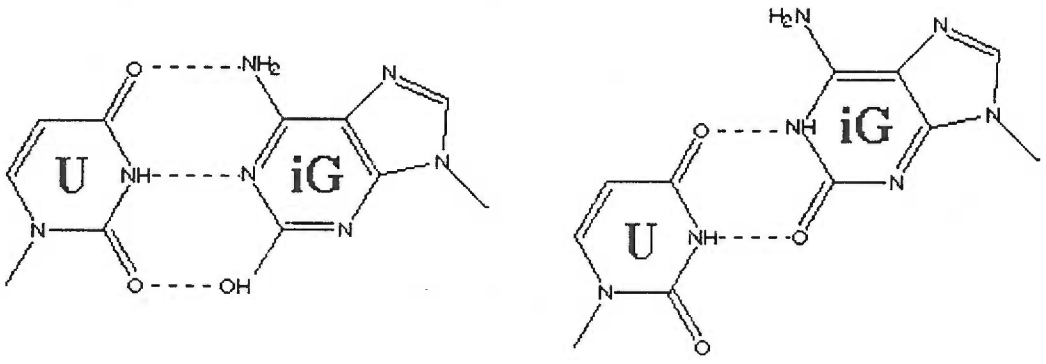


Figure 5.3. Two possible configurations of U·iG (uracil·isoguanine) base-pairs with iG in different tautomeric forms (Robinson et al., 1998). The Watson-Crick base-pair (iG is in enol form) in shown on the left and the wobble base-pair (iG is in keto form) is shown on the right.

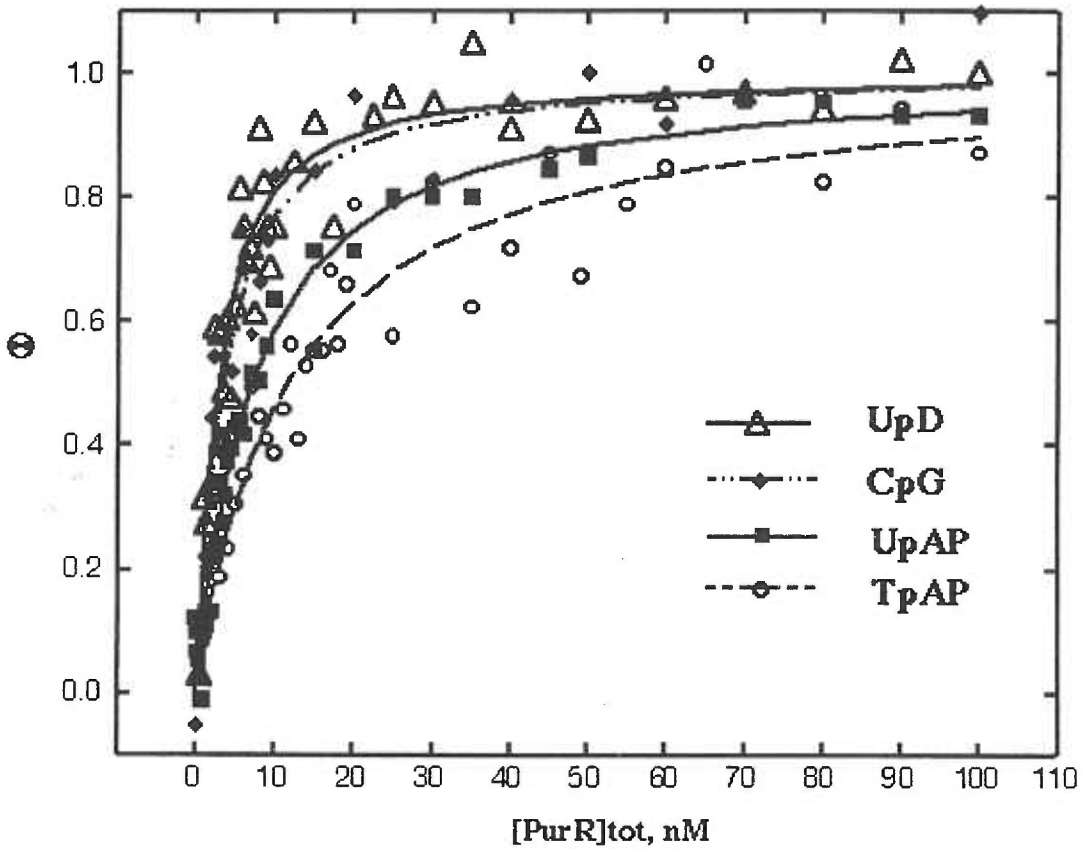


Figure 5.4. Binding of PurR to native *purF* operator (CpG) as well as modified UpD, UpAP, and TpAP operators, in the presence of hypoxanthine. The binding data (fractional saturation, Θ, of the DNA operator with PurR) are plotted as a function of total PurR dimer concentration. All data are fitted by nonlinear least-squares regression to the equation describing the simple binding model.

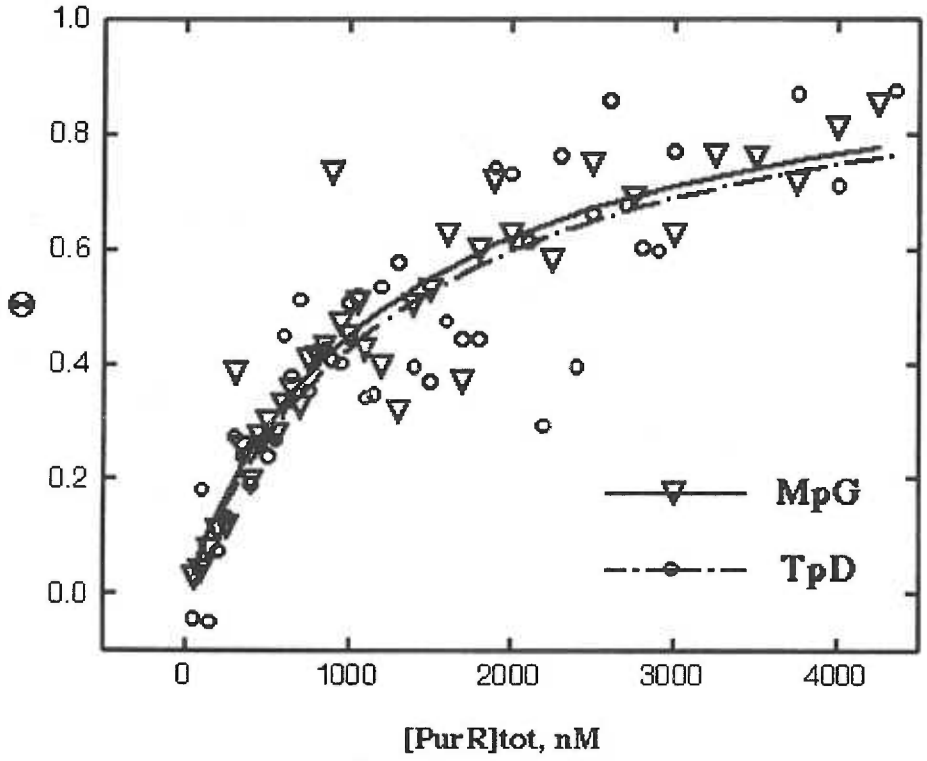


Figure 5.5. Binding of PurR to modified 5meCpG and TpD operators in the presence of hypoxanthine. The binding data (fractional saturation, Θ, of the DNA operator with PurR) are plotted as a function of total PurR dimer concentration. All data are fitted by nonlinear least-squares regression to the equation describing the simple binding model.

Figure 5.6. A close view of the central CpG step in the *purF* operator [Schumacher et al., 1994] with a modelled 5-methyl group. For clarity, only the A8-C9-G9'/C9-G9'-T8'bases are shown and labelled in green. The DNA atoms are coloured by atom type where carbons are white, oxygens are red, nitrogens are blue, and phosphors are yellow. Carbons of modelled the 5-methyl groups are shown in green. The side chains of residues Leu54 and Leu54' are shown in pink. Distances are shown in Å as yellow dashed lines.

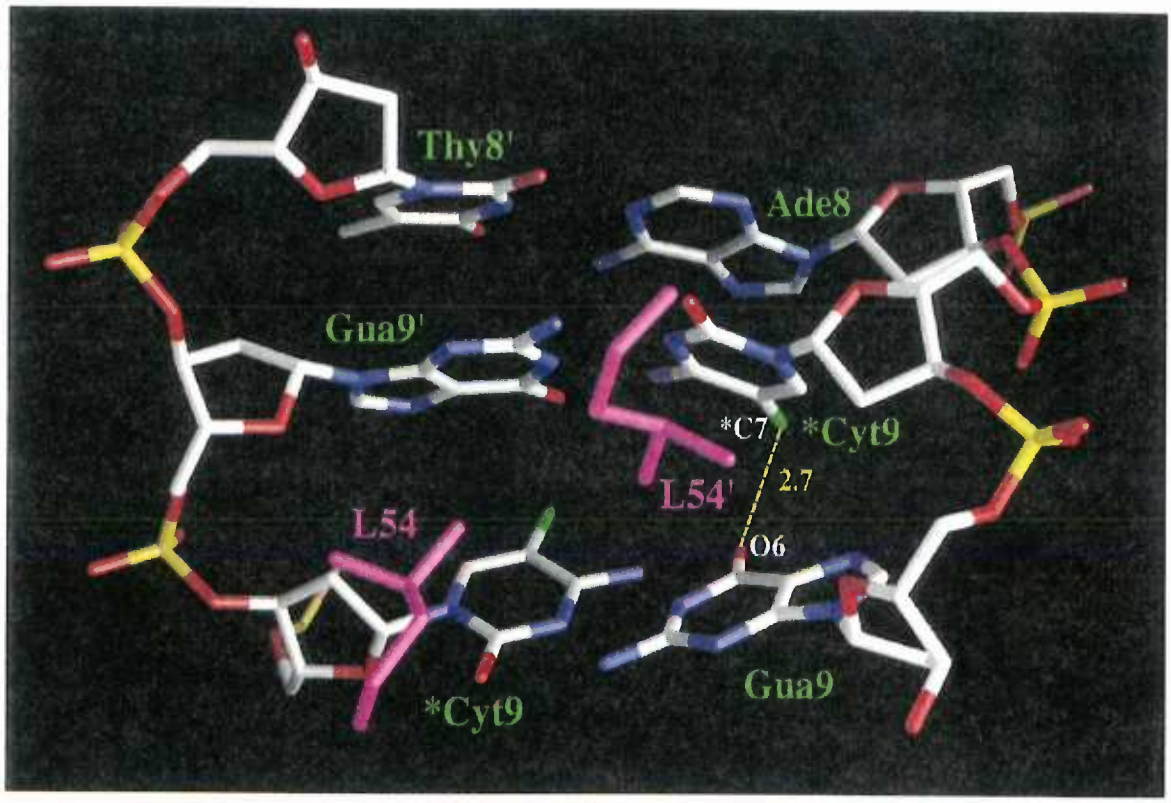

Thy8'
Ade8
Gua9'
*C7'
*Cyt9
2.7
L54'
L54
O6
*Cyt9
Gua9

Figure 5.7. Structure of the PurR in complex with hypoxanthine and with a modified *purF* operator containing the central TpD step. Labelled are the corepressor-binding domain (CBD) and the DNA-binding domain (DBD). The CBD is coloured purple, the helix-turn-helix-loop-helix motif is coloured pink, and the hinge helix is coloured green. The hypoxanthine corepressor is shown as yellow balls and sticks, and the intercalating residues Leu54 and Leu54' are shown as black balls and sticks. A red arrow indicates the location of the central TpD step.

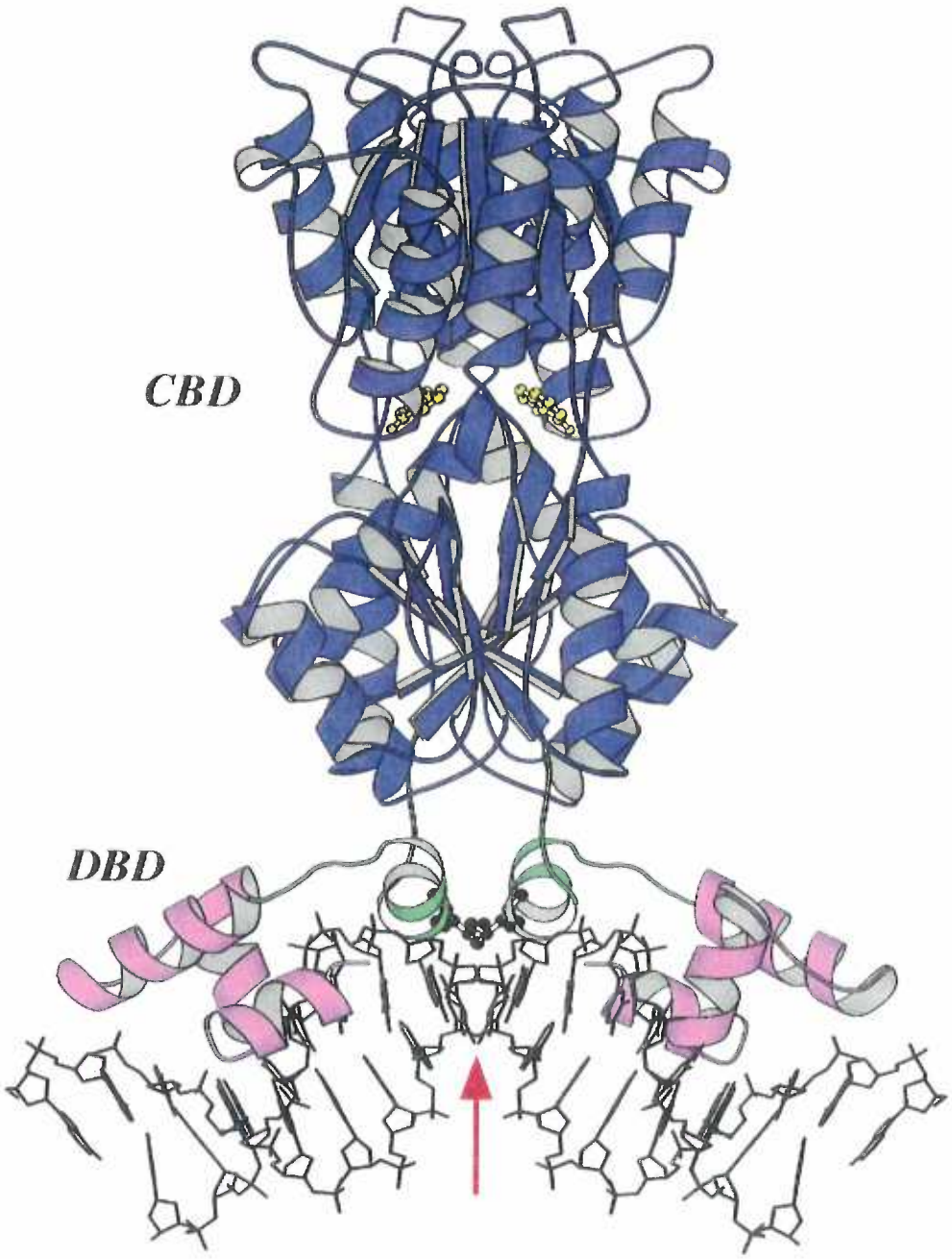

CBD
DBD

Figure 5.8. Overlay of the native and four modified *purF* operators. Native *purF* operator DNA is shown in blue, TpD operator is shown in black, UpAP operator is shown in red, UpD operator is shown in cyan, and 5meCpG operator is shown in green. For clarity, only the DNA-binding domain of the TpD complex is shown. Hinge helices are shown in green, and helix-turn-helix-loop-helix (HTHLH) domains are shown in pink. Intercalating residues Leu54 and Leu54' are shown as black balls and sticks. An arrow points at the location of the central base-pair step.

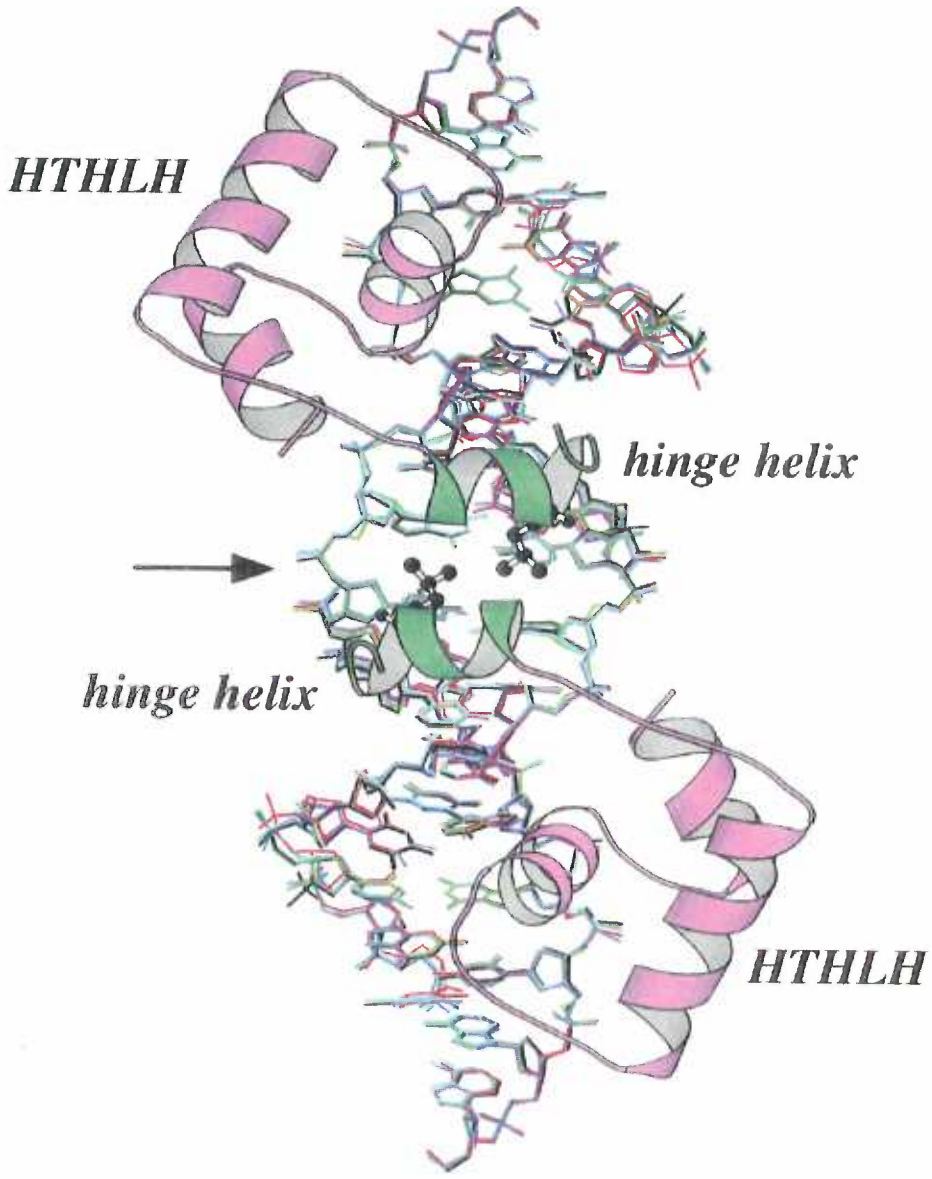

HTHLH
hinge helix
hinge helix
HTHLH

Figure 5.9. A close view of the central sequences in the native and modified *purF* operators. DNA bases are labelled in green. The DNA atoms are coloured by atom type where carbons are white, oxygens are red, nitrogens are blue, and phosphors are yellow. Side chains of residues Leu54 and Leu54' are shown in pink. Distances are shown in Å as yellow dashed lines.

A. CpG operator (native).

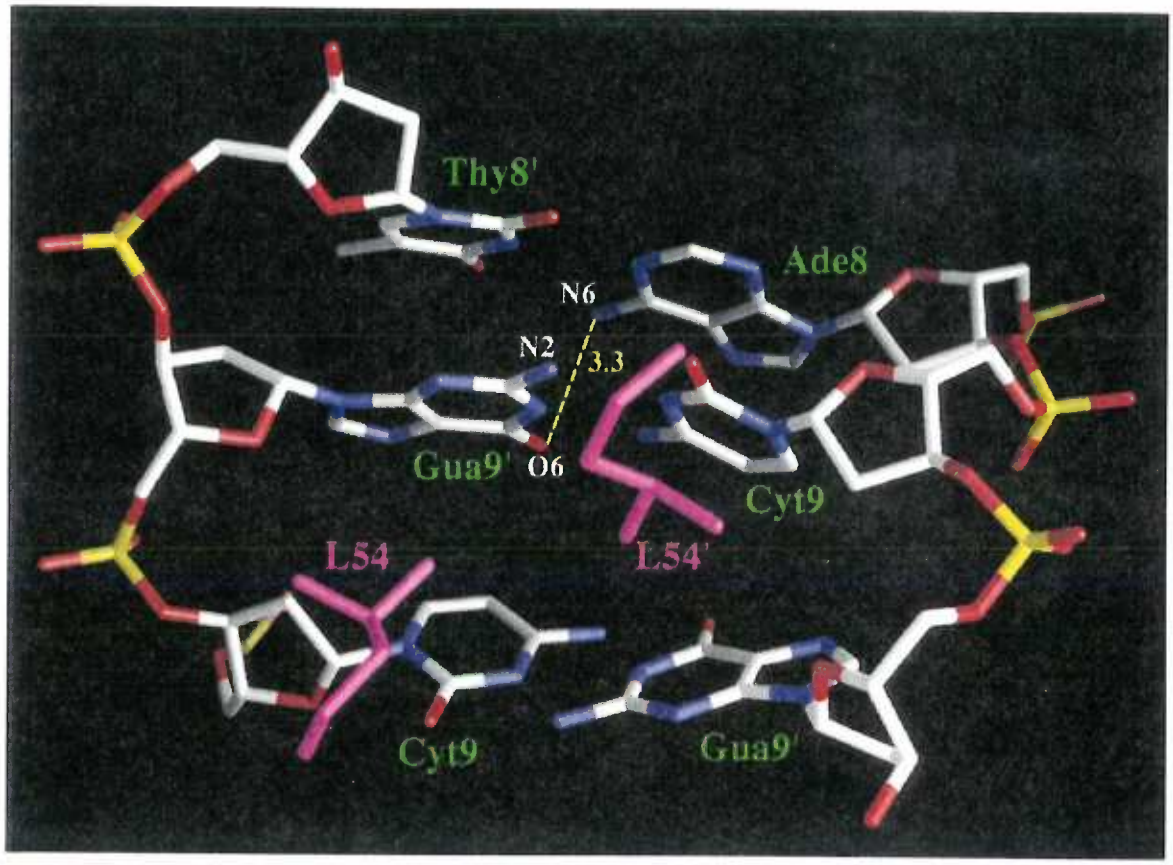
B. UpD operator.

C. 5meCpG operator.

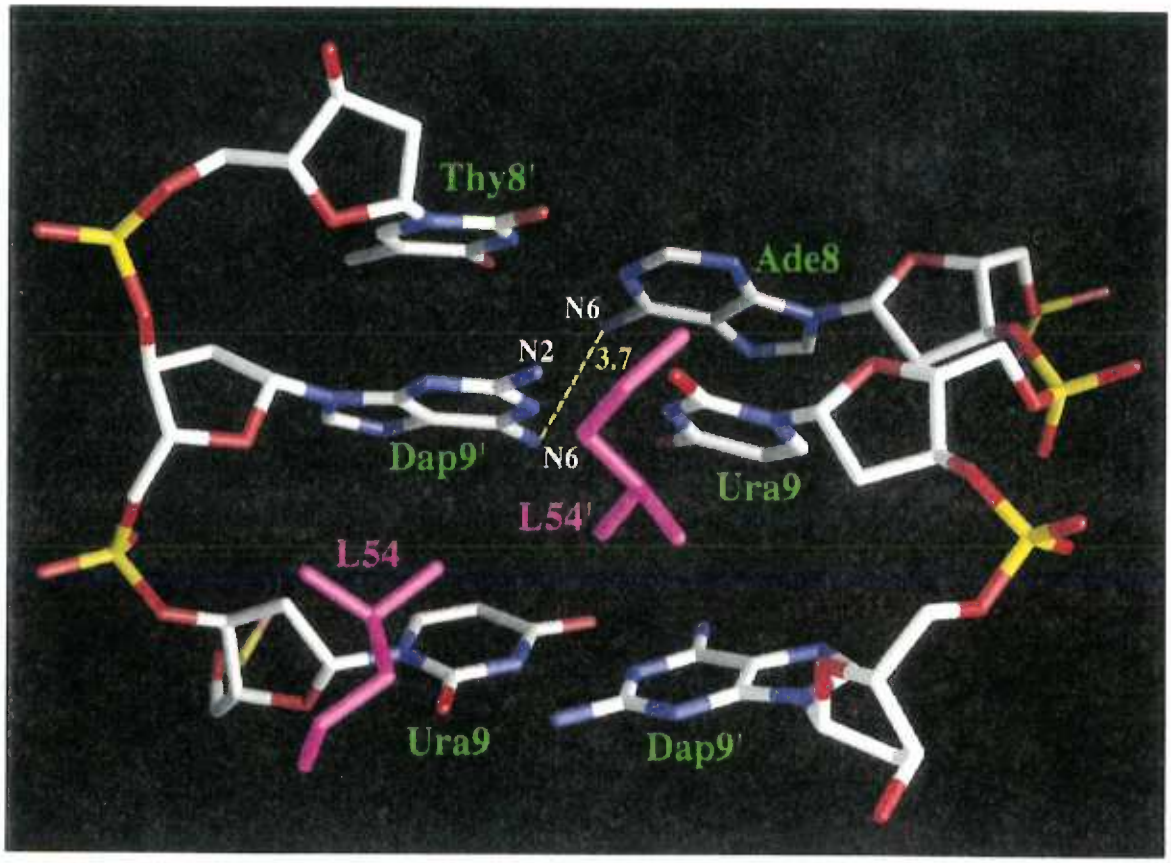
D. TpD operator.

E. UpAP operator.

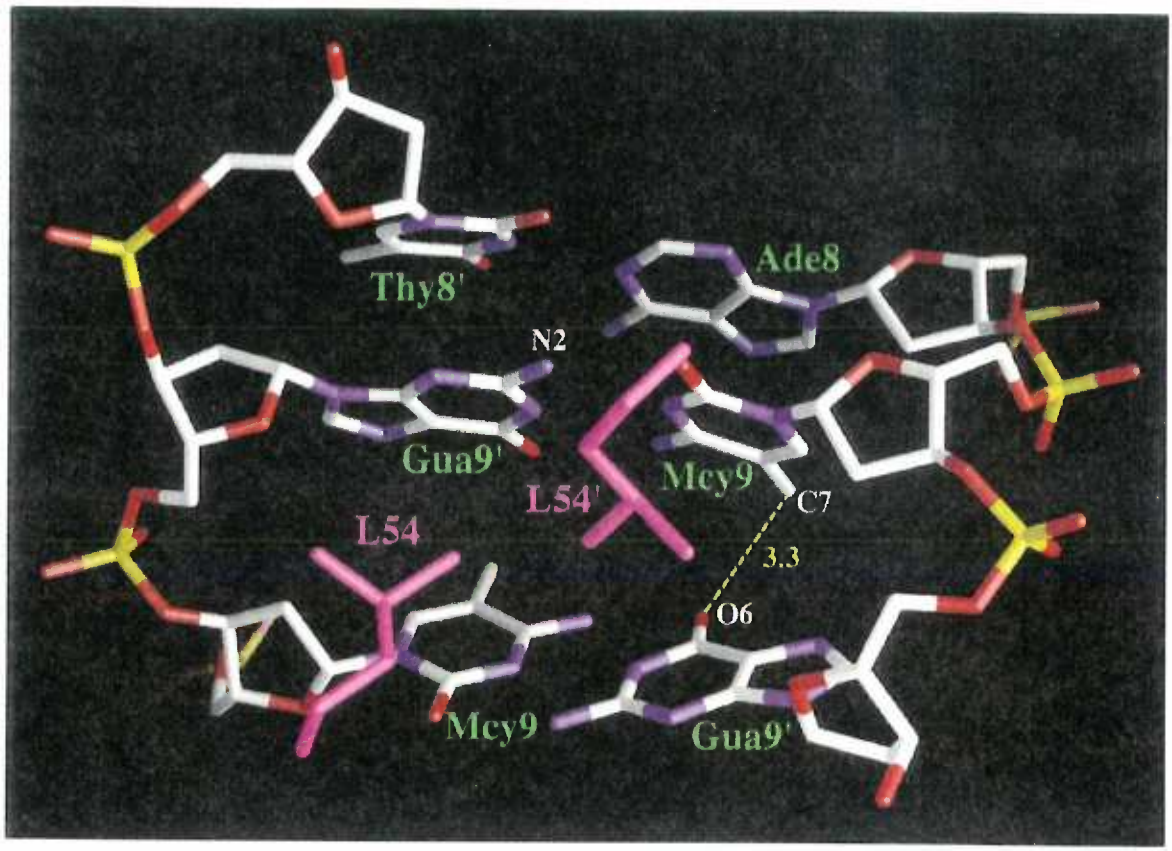
A



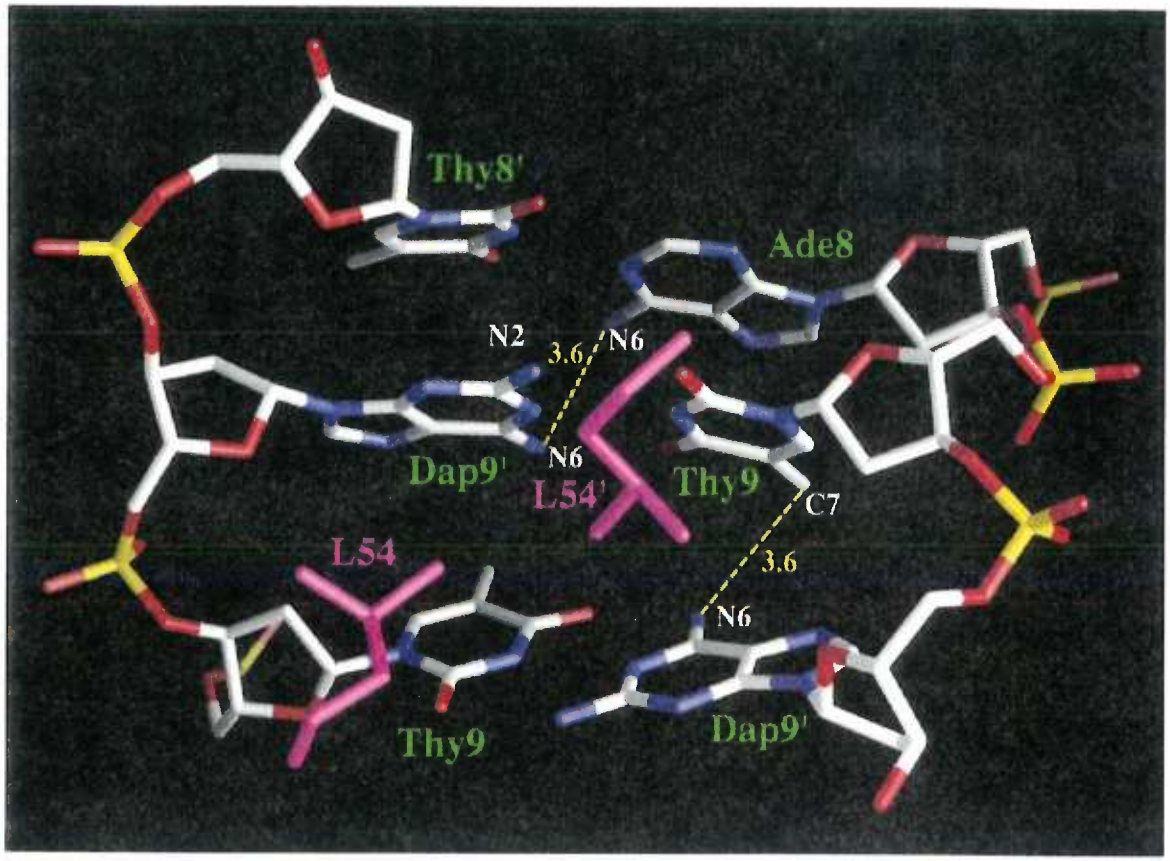
B



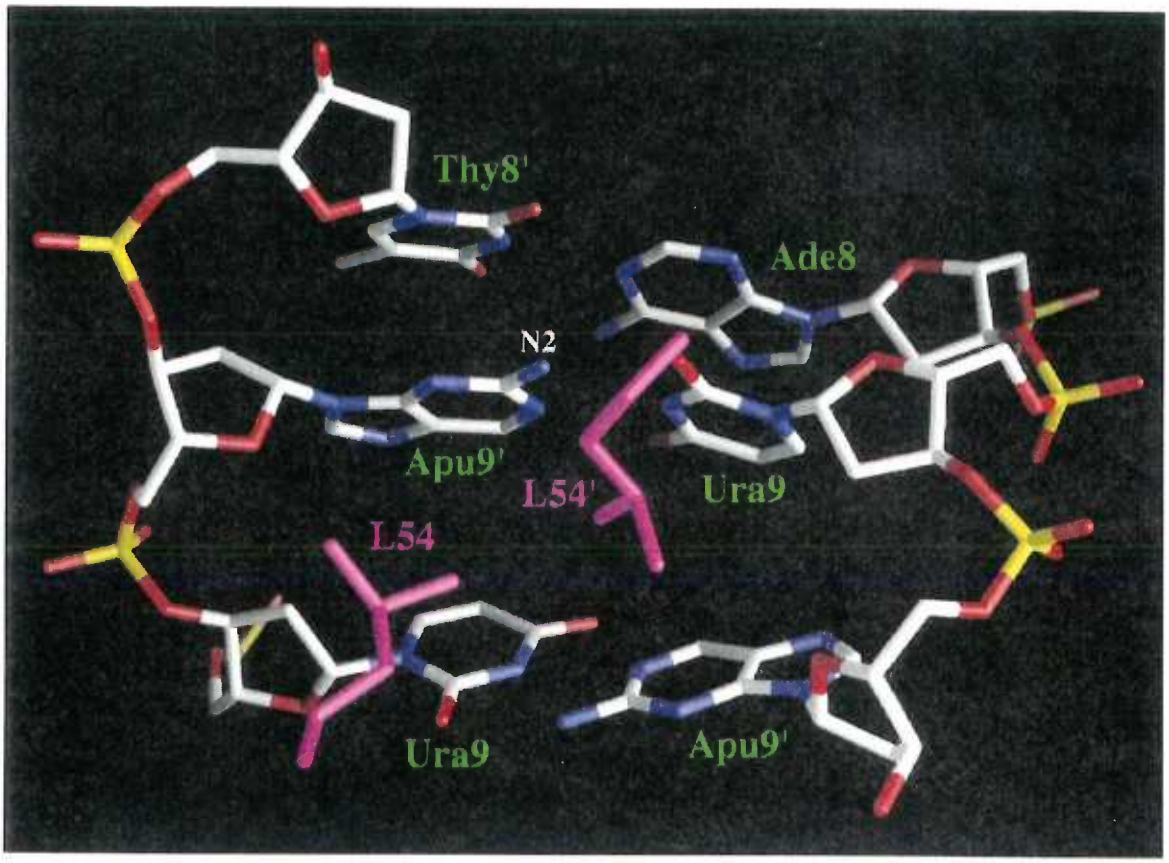
C



D



E



CHAPTER 6

SUMMARY AND CONCLUSIONS

BmrR project.

Multidrug-binding proteins, including transporters and transcriptional regulators, demonstrate an unusual ability to recognize structurally and chemically diverse toxic compounds, which contrasts dramatically with the narrow chemical specificity of the vast majority of proteins. In addition, these proteins have gained considerable attention of the biomedical community because of their ability to protect tumor cells and pathogenic microorganisms from antitumor and antimicrobial agents. Yet, despite extensive biochemical studies, the structural mechanism of multidrug recognition remains obscure, primarily because of the dearth of structural information on any of the multidrug-binding proteins.

To obtain insight into the enigmatic mechanism of multidrug recognition, we focused on the multidrug-binding transcription regulator BmrR from *Bacillus subtilis*, which controls the expression of the multidrug transporter gene, *bmr*. Specifically, we undertook crystallographic studies on the carboxy-terminal drug-binding/dimerization domain of BmrR (BRC), which includes residues 121-279. I initially solved the crystal structure of the apo BRC to 2.7 Å resolution by multiple isomorphous replacement and subsequently solved the structure of BRC in complex with a drug, tetraphenylphosphonium (TPP), to 2.8 Å resolution by molecular replacement.

The structures revealed a unique fold of the BRC molecule, reminiscent of the β-

barrel motif taken by lipocalin proteins, and showed that BRC exists as a metal-dependent dimer. More importantly, the structures brought to light for the first time the architecture of a binding site capable of binding multiple structurally dissimilar compounds.

Specifically, the binding of cationic TPP is mediated by stacking and van der Waals contacts with multiple hydrophobic and aromatic residues of the binding pocket and by an electrostatic interaction with the carboxylate group of the buried glutamate residue Glu134, an interaction favoured by the low dielectric constant of the protein interior. In the apo structure, this binding site is masked by an α-helix and thus is rendered completely inaccessible. However, in the complex, the helical 'seal' of the binding site unfolds and relocates away from the protein thus allowing a ligand to take its place. The ability of this helix to undergo a conformational transition, perhaps, even in the absence of drug, is indicated by its average temperature factor, which is significantly higher than the average temperature factor for the rest of the protein and by its sensitivity to trypsin.

The interactions between BRC and tetraphenylphosphonium suggest a similar binding mode for other hydrophobic cationic ligands of BmrR. Indeed, when rhodamine was manually docked in the binding pocket of the template BRC-TPP structure, this high-affinity drug appeared to make van der Waals and stacking contacts similar to those observed in the BRC-TPP complex. More importantly, the positive charge on rhodamine seem to approach closer the carboxylate of Glu134, which likely accounts for most of its ~100-fold higher binding affinity in comparison with TPP. Furthermore, the preliminary

structure of the BRC bound to compound II (in progress) shows that compound II binds in the same location and almost in the same orientation as TPP.

Taken together, these crystallographic and modelling studies propose that residue Glu134 is the key to cation selectivity while a number of nonpolar and aromatic side chains impose specific requirements on drug size and shape. Now, we would like to see whether other compounds, in particular high-affinity ligands rhodamine, crystal violet, and compound IV indeed can bind to the same site and whether their high affinity arises from proximity of their positive charges and the negatively charged carboxylate moiety of the glutamate.

From our studies, a likely principle of multidrug recognition based on electrostatic interactions has emerged. Since many mdr transporters have negatively charged amino acid residues (Asp and Glu) in their transmembrane domains, we hypothesized that other multidrug binding proteins, including multidrug transporters, may use the same principle to recognize their cationic substrates.

Finally, the mechanism of DNA-binding and the allosteric mechanism of multidrug-induced transcription activation of the *bmr* gene by BmrR remain enigmatic. Clearly, the determination of the molecular mechanism by which drug-binding turns BmrR into a transcription activator entails further biochemical investigation and, ultimately, the structure of a BmrR-promoter-drug ternary complex as well as BmrR-promoter binary complex.

PurR project.

Direct recognition of DNA bases is the principal mode of sequence-specific

recognition by many DNA-binding proteins. However, DNA can indirectly affect this recognition. We employ the Purine Repressor from *E.coli* as a model system to explore the specificity of DNA-binding. In its sequence-specific recognition, PurR relies primarily on the direct base-specific contacts; however, it appears to read the conserved central CpG sequence indirectly. In particular, we addressed the question of why PurR discriminates against the central TpA sequence and other base-pair steps in absolute favour for CpG.

Initially, we compared chemical composition of CpG and TpA and dissected the chemical differences between them with particular emphasis on exocyclic groups. To determine the role of these exocyclic groups in PurR specificity, we designed a series of *purF* operators in which the central CpG sequence was modified to CpI, TpA, UpD, TpD, UpAP, TpAP, 5meCpG, and UpiG. We then employed two different strategies. First, we measured the binding affinities for modified operators containing these altered central base-pair step. By comparing them with the binding affinity for the native operator, we estimated the thermodynamic contribution of each individual group to binding in terms of $\Delta\Delta G$. Second, we solved crystal structures of PurR in complex with different modified operators. This integrated approach has allowed us to correlate the thermodynamic effect of these modifications with structural changes.

We identified the key determinants of PurR specificity for the central sequence: a 5-methyl group on the pyrimidine and a 2-amino group on the purine. PurR prefers the CpG sequence because it has the 2-amino group on guanine and lacks a 5-methyl group on cytosine. The TpA step has the exactly opposite situation: it lacks a 2-amino group on adenine and has the 5-methyl group on thymine. The 2-amino group on the purine base

is absolutely required for specific binding. In its absence, the specific PurR-operator complex cannot form. We hypothesize that the 2-amino group widens the minor groove, which provides a template for the folding or stabilization of the hinge-helix and thus, allows the formation of the high-affinity sequence-specific PurR-DNA complex. However, it remains to be shown whether any group of a similar size could functionally replace the 2-amino group.

The bulk of the 5-methyl group on the pyrimidine provides a steric barrier to the compression of the major groove that occurs when PurR kinks the DNA. Specifically, such compression brings the 5-methyl group into unfavourable proximity to the group at position 6 of a neighbouring purine from the central step. The PurR-DNA complex can still form, but additional energy is required to overcome this steric barrier and attain the proper global kinked DNA conformation. Consequently, one of these two positions must be vacant to allow high-affinity binding.

Overall, our results suggest that the geometry of PurR-operator complex imposes a set of specific requirements for exocyclic groups in the major and in the minor grooves of the central YpR step. The specificity of PurR for the central CpG group stems from the ability of the CpG operator DNA to assume a particular bent conformation and to facilitate a coil-to-helix transition in the hinge helix.

We would like to determine the affinity of a mutant PurR protein with a stabilized hinge-helix for both the wild-type *purF* operator and for the CpI and TpA substituted operators. These binding studies would be paralleled by proteolysis and structural studies.

APPENDIX A

A POSSIBLE MECHANISM TO EXPLAIN THE POLYSUBSTRATE SPECIFICITY OF MULTIDRUG-EFFLUX TRANSPORTERS: LESSONS FROM THE STRUCTURE OF A SOLUBLE MULTIDRUG-RECOGNIZING PROTEIN

Multidrug resistance transporters (mdrs) are membrane proteins that are found in nearly every cell. These proteins bind multiple structurally and chemically dissimilar compounds and remove them from the cell in an ATP or proton motor force-dependent manner. First discovered in mammalian cells (P-glycoprotein) [Gottesman and Pastan, 1993; Sharom, 1997], multidrug transporters were later identified in bacteria [Nikaido, 1994; Paulsen et al., 1996a] and yeast [Goffeau et al., 1997]. Only Mycoplasma appears to lack these proteins [Fraser et al., 1995]. At present, the number of identified multidrug transporters exceeds one hundred with many cells expressing several transporters. Known multidrug transporters belong to at least four different protein superfamilies [Gottesman and Pastan, 1993; Paulsen et al., 1996a]. However, even within a superfamily, individual multidrug transporters have a higher sequence similarity to substrate-specific transporters than to each other.

These mdr proteins have come to the forefront of the biomedical research because of their ability to protect tumor cells and pathogenic microorganisms from antitumor and antimicrobial agents. However, they are equally interesting from a purely theoretical standpoint. Indeed, the drug substrates of multidrug transporters share no structural similarity, except that the majority of them are lipophilic molecules carrying a positive electrostatic charge (Figure 1). Moreover, multidrug transporters interact with a wide range of inhibitors, which, presumably by a competitive mechanism, prevent transporters from pumping drugs out of the cell. Like substrates, these inhibitors have no common structural motifs that would provide specificity. How then can a single protein recognize a number of remarkably diverse substrates and inhibitors?

Several mechanisms have been posited to explain the ligand-recognition and efflux properties of multidrug transporters. One model suggests that multidrug transporters do not bind the transported drugs directly but rather indirectly promote the efflux of diverse positively charged compounds through dissipating the membrane potential [Roepe, 1995]. Another hypothesis is based on the accumulating evidence [Homolya et al., 1993; Bolhuis

et al., 1996; Shapiro et al., 1997] that many, if not all, substrates of multidrug transporters are expelled not from the cytoplasm, but directly from the cell membrane [Higgins and Gottesman, 1992]. Such mode of action has been described as a "hydrophobic vacuum-cleaner" [Gottesman and Pastan, 1993]. Indeed, it is possible to imagine a multidrug transporter perpetually purging the membrane and indiscriminately expelling all the molecules loosely associated with it, including all the moderately hydrophobic drugs.

However, the existing evidence strongly indicates that multidrug transporters interact directly with diverse transported drugs and inhibitors. First, structural alterations of the transporter substrates dramatically affect their transport affinities [Tang-Wai et al., 1993]. In fact, multidrug transporters are polyspecific rather then nonspecific, which means that they promote the efflux of many different compounds but not just any lipophilic compound. Second, a large number of mutants of P-glycoprotein [Germann, 1996; Hanna et al., 1996; Sharom, 1997] and several other multidrug transporters [Ahmed et al., 1993; Paulsen et al., 1996b; Paulsen et al., 1996c; Kiyachko et al., 1997; Edgar and Bibi, 1999] exhibit altered transport properties and affinities for specific drug substrates and specific inhibitors. Finally, multiple photoreactive substrates and inhibitors of P-glycoprotein have been shown to bind directly to P-glycoprotein upon irradiation [Cornwell et al., 1986; Greenberger et al., 1991; Morris et al., 1991; Bruggemann et al., 1992] whereas other substrates and inhibitors compete for this binding. Thus, the specific structure of the transporter along with the properties of substrates themselves dictates the spectrum of substrates and inhibitors. It appears, therefore, that multidrug transporters conform to the traditional view on the mechanism of membrane transport, whereby the molecule of a substrate first interacts with some kind of a substrate-recognition site within a transporter and only then is effluxed from the cell. The only puzzling difference from regular substrate-specific transporters is that the recognition site of a multidrug transporter can interact, in an unexplained way, with a large array of dissimilar molecules.

The problem of multidrug recognition has been addressed by multiple experimental approaches, yet our understanding of the molecular interactions is still in its infancy. Formally, the nature of these interactions can be reliably explored only by analyzing the structures of the transporter molecule and of its complexes with substrates and inhibitors, at the atomic level. However, the structural analysis of membrane-associated multidrug transporters presents serious technical difficulties and has not yet yielded an atomic resolution structure.

To circumvent the technical problems associated with the crystallization and structural analysis of membrane proteins, we chose to focus on a soluble protein that also displays the ability to recognize multiple drugs. This protein, BmrR from *Bacillus subtilis*, is a MerR-family transcription regulator [Summers, 1992; Hidalgo et al., 1997] that controls the expression of the multidrug transporter gene, *bmr* [Ahmed et al., 1994]. BmrR mediates induction of *bmr* expression by a number of structurally diverse hydrophobic cations, such as rhodamine and tetraphenylphosphonium (Figure 1), many of which are also substrates of the Bmr transporter [Markham et al., 1997]. BmrR consists of the N-terminal DNA-binding domain that binds to the *bmr* promoter, a linker region, and an inducer-binding/dimerization domain, designated BRC (for BmrR C-terminus). The individually expressed BRC, a 159-residue protein, retains the ability to dimerize and to bind inducers with the same affinities as the full-length BmrR [Markham et al., 1996].

The recent crystallographic analysis of apo BRC and of its complex with one of the inducers (drugs), tetraphenylphosphonium (TPP), revealed the following mechanism of ligand binding [Zheleznova et al., 1999]. TPP penetrates into the hydrophobic core of the protein, where it forms a number of van der Waals and stacking interactions with the surrounding hydrophobic residues, but most importantly, an electrostatic interaction with a buried glutamate residue, Glu134 (Figure 2B). Inside the hydrophobic binding pocket, this glutamate provides the negative charge necessary to attract and bind a positively charged drug. The low dielectric constant of the protein interior (~2) would strongly

favour this electrostatic interaction. Moreover, this interaction, between the positively charged ligand and negatively charged glutamate, is the key to cation selectivity of the BRC protein. Even the conservative substitution of glutamate to the isosteric glutamine abolishes drug binding without perturbing the structural integrity of the protein.

To allow the ligand to gain access to this glutamate of BRC, a short amphipathic α-helix, which forms a boundary between the hydrophobic core and the surrounding medium (Figure 2A), undergoes a dramatic conformational change: it unwinds and relocates away from the protein thus allowing a ligand to take its place. In the absence of a drug, this amphipathic helix serves as a "lid" for the drug-binding site, rendering it inaccessible to solvent. The ability of this helix to undergo a conformational transition, perhaps, even in the absence of drug, is indicated by its average temperature factor, which is significantly higher than the average temperature factor for the rest of the protein. Furthermore, the helix is protease sensitive, i.e., it can be cleaved by trypsin almost in its middle in the absence and, very readily, in the presence of drug. Crystallographic analysis of the BRC-TPP complex does not reveal any electron density for the unwound helix, which indicates that it fluctuates between several locations. For stereochemical reasons, the unwound helix cannot move very far away into the solvent and is likely to form an outside steric barrier that hinders the dissociation of the ligand. Also, the hydrophobic residues of the unwound helix may form additional transient contacts with the bound ligand.

Initial crystallographic analyses of the BRC complexes with its other ligands (in progress), have demonstrated that they indeed bind similarly, if not identically, to the protein. Therefore, one key to the unusual ligand promiscuity of BmrR is the electrostatic interaction between positively charged ligands and the negatively charged glutamic acid in the binding pocket. Moreover, structural analysis and modelling of different BRC-ligand complexes suggests that the closer the positive charge of the ligand is to the Glu134 carboxylate, the higher the binding affinity. The proximity of the charged moieties depends not only upon the structure of the ligand but also on the architecture of the binding site.

Namely, hydrophobic and aromatic amino acids in the binding site are arranged in a way that precludes the binding of every positively charged drug. However, at the moment, the degree of flexibility of the binding site is unknown. The hydrophobic side chains may rotate to accommodate drugs with different geometry, or they may remain in relatively fixed orientations.

It is tempting to speculate that multidrug transporters employ a similar mechanism based on electrostatic and steric complementarity between their substrates and their binding sites. As an example, we consider a proton-dependent multidrug transporter that uses a transmembrane proton gradient, or proton-motor force (PMF), to expel drugs from the cell. PMF-dependent multidrug transporters preferentially extrude hydrophobic cations and are thought to function as multidrug/proton antiporters [Paulsen et al., 1996b]. As suggested by drug-binding data and in a manner analogous to BmrR, the physical characteristics of a compound rather than its exact structure appear to be key determinants in the substrate selectivity of multidrug transporters, analogously to BmrR. For example, the *E.coli* multidrug transporter, EmrAB, exports hydrophobic quinolones but not their hydrophilic analogues [Lomovskaya and Lewis, 1992]. As another example, all lipophilic substrates of the QacA/QacB efflux system of *S. aureus* contain positively charged groups. Interestingly, the difference in relative selectivities depends on the number of positively charged groups in a substrate: QacA transports divalent and monovalent cations whereas QacB recognizes only monovalent cations [Paulsen et al., 1996c]. Thus, the cationic nature of multidrug-transporter substrates indicates that electrostatic interactions are critical in multidrug binding.

We would like to propose the following model for multidrug binding by multidrug transporters. First, drugs are detected primarily in the plasma membrane. The relative hydrophobicity of drugs suggests that they would partition into the lipid phase prior to binding a multidrug transporter, and indeed, multiple data imply that mdr transporters remove their substrates directly from the membrane [Homolya et al., 1993; Bolhuis et al.,

1996; Shapiro et al., 1997]. Second, to attract and to bind the cationic drug, the internal binding site of the transporter must have an electronegative residue(s) complementing the positive charge of a drug. The compounds then gain access to the substrate-binding site of the transporter through a transient opening made by a flexible secondary structure element, most likely, an α-helix *à la* BRC. The shape of this entryway will promote substrate discrimination against sterically unsuitable compounds.

In support of the second requirement, electronegative residues (aspartates and glutamates) have been found in predicted as well as experimentally identified membrane-spanning regions of most proton-dependent multidrug transporters [Paulsen et al., 1996a; others]. Moreover, mutational analysis of several multidrug transporters has shown that some acidic residues are critical to binding and essential to multidrug transport. Specifically, a glutamate in the first transmembrane domain is conserved among all multidrug transporters of the small multidrug resistance (SMR) family. This glutamate E13 is essential for transport activity of Smr protein, because substitution to any other residues (even aspartate) abolish transport [Paulsen et al., 1996b]. Similarly, in the *S. aureus* multidrug transporter QacA, a transmembrane acidic residue at position 322 or 323 is required for substrate binding [Paulsen et al., 1996c]. Finally, the *E. coli* multidrug resistance protein, MdfA has only one membrane-embedded glutamate at position 26, in the middle of putative transmembrane segment 1 [Edgar and Bibi, 1997]. More importantly, the mutations at position 26 have a drastic effect on the substrate recognition profile of MdfA in general suggesting that MdfA causes multidrug resistance by directly interacting with and transporting the drugs. Specifically, replacing E26 with the positively charged residue lysine abolishes the multidrug resistance activity against positively charged drugs, but retains chloramphenicol efflux and resistance. In contrast, when the negative charge is preserved in a mutant with aspartate instead of E26, chloramphenicol recognition and transport are drastically inhibited; however, the mutant exhibits almost wild-type multidrug resistance activity against lipophilic cations [Edgar and Bibi, 1999]. These results enabled

clear distinction between transport activity and substrate recognition, and indicated that although the negative charge at position 26 is not essential for active transport, it dictates the multidrug resistance character of MdfA.

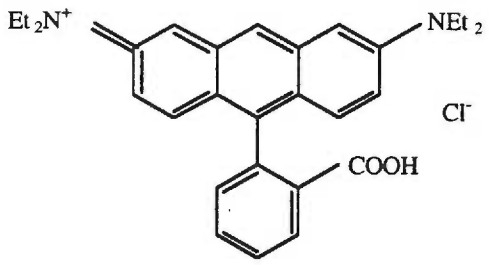
The electrostatic attraction between the negative charge on the transporter and the positively charged drug would be extremely favourable because of the low dielectric constant of the membrane interior. The strength of an electrostatic interaction is diminished significantly in an aqueous environment because of the competing interactions between the charged molecules and waters. Moreover, in the presence of salts, ion screening significantly reduces interactions between charges. By contrast, in the hydrophobic, ion-free environment of the membrane, electrostatic interactions are very potent. We also posit that the low dielectric constant of the membrane interior (~2 to 4 versus ~78 for water) might magnify the effective electrostatic field of glutamate side chain, and as the result, charges can "sense" one another at distances 6-fold longer than in water. Therefore, a negative charge in the binding site of a multidrug transporter could act like a magnet, effectively attracting and ultimately sequestering any appropriately-shaped hydrophobic cation from the lipid bilayer.

Clearly, an exposed, uncomplemented negative charge is thermodynamically unfavourable in the hydrophobic membrane environment. In our model, the binding site is shielded but can be readily accessed through a gap created by a thermodynamically unstable secondary-structure element, like in BRC. Indeed, it is possible to imagine a flexible α-helix which covers the entry port to the intramembrane acidic residue of the transporter, constantly undergoing a helix-to-coil transition. Sequence analysis of multidrug transporters reveals abundant proline and glycine residues that are known to destabilize α-helices [Branden and Tooze, 1998].

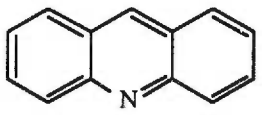
In summary, we have proposed a mechanism of multidrug recognition that accommodates the existing mechanistic theories and is in accord with accumulated

biochemical, mutational, and structural data. A hydrophobic cationic drug can easily insert into the membrane lipid phase and like any compound is partitioned between the hydrophilic cytosol and the hydrophobic inner leaflet of the membrane. Once inside the membrane, the drug will be attracted to the transiently-exposed negatively-charged binding site and can get access to it through a secondary-structure element, most probably an α-helix, that undergoes a helix-coil transition as seen in BmrR. As the final step in drug extrusion, we propose a second conformational change in the multidrug transporter that links proton import to drug export. Such a mechanism explains how a transporter can recognize structurally diverse substrates and yet exhibit certain selectivity. The geometry of the binding site does not allow just any hydrophobic cation to bind, and any sterically inappropriate compound would stall at the portal and ultimately diffuse away. Indeed, in multidrug transporters, a spectrum of substrate affinities can be affected by mutations changing the geometry or shape of the binding site. In addition, proton binding may participate in selection of appropriately shaped ligands as well as in neutralizing the negative charge in the binding site.

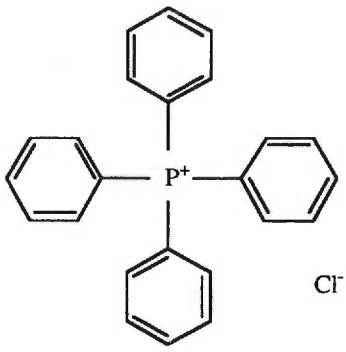
In conclusion, the structures of the apo and drug-bound BRC revealed that the broad specificity of ligand binding arise from the simple electrostatic and steric complementarity of a drug and a multidrug-binding protein. Whether the same principle works in multidrug transporters can be explicitly verified only upon their structural analysis. Perhaps, the most important lesson of the BRC structure is that the polyspecificity of ligand recognition does not require any special mechanisms lying beyond the realm of existing knowledge. This alone significantly reduces the flair of mystery surrounding the phenomenon of multidrug transporters.



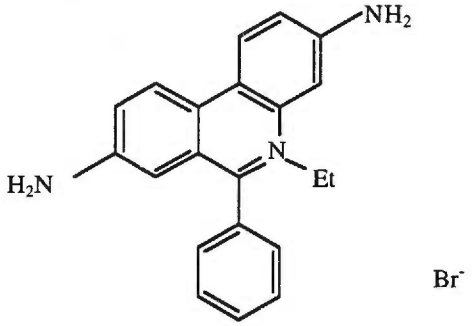
Rhodamine



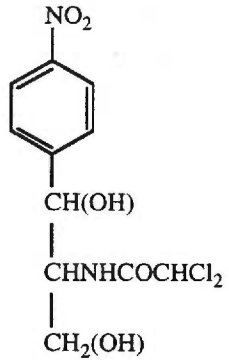
Acridine



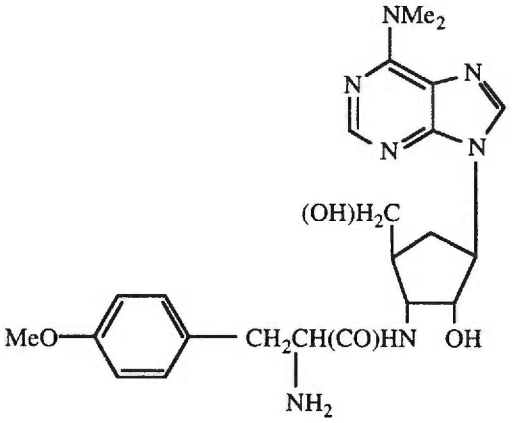
Tetraphenylphosphonium



Ethidium bromide



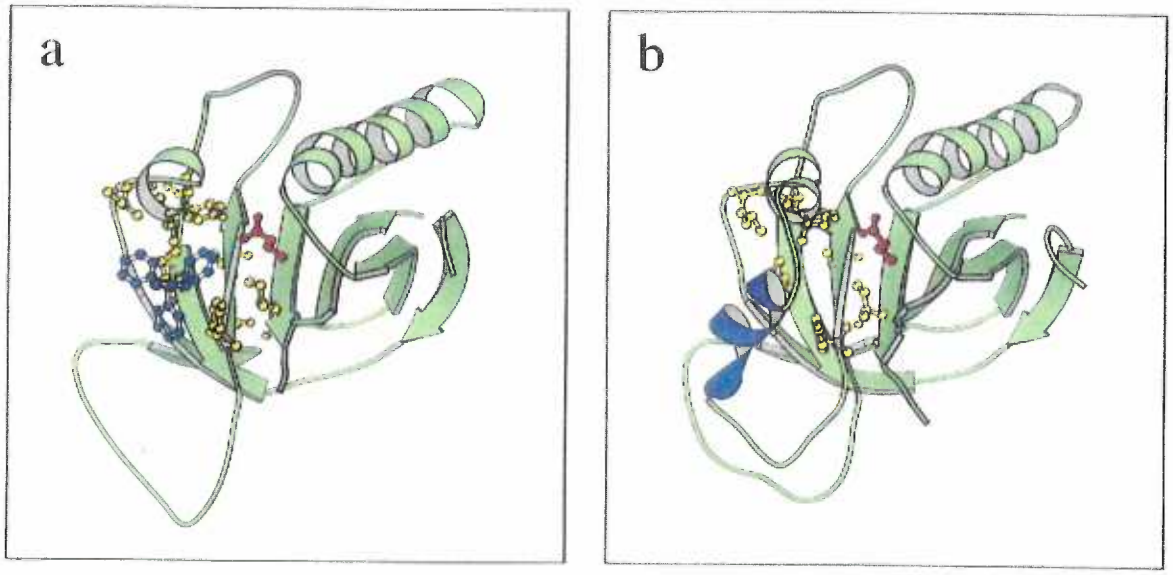
Chloramphenicol



Puromycin

Figure 1. Selected substrates of multidrug transporters.

Figure 2. Structures of BRC in complex with TPP (A) and in apo form (B) are shown as green ribbons. Hydrophobic and aromatic residues composing the drug-binding pocket are shown as yellow balls and sticks. The glutamate E134 is shown as red balls and sticks. The bound TPP molecule (A) is shown as blue balls and white sticks. An α-helix in the apo form that unfolds to allow drug binding is shown as a blue ribbon (B).


a
b

LITERATURE CITED

Abergel, C., Moulard, M., Moreau, H., Loret, E., Cambillau, C., and Fontecilla-Camps, J.C. (1991). Systematic use of the incomplete factorial approach in the design of protein crystallization experiments. J. Biol. Chem. *266*, 20131-20138.

Ahmed, M., Borsch, C.M., Neyfakh, A.A., and Schuldiner S. (1993). Mutants of the *Bacillus subtilis* multidrug transporter Bmr with altered sensitivity to the antihypertensive alkaloid reserpine. J. Biol. Chem. *268*, 11086-11089.

Ahmed, M., Borsch, C.M., Taylor, S.S., Vazques-Laslop, N., and Neyfakh, A.A. (1994). A protein that activates expression of a multidrug efflux transporter upon binding the transporter substrates. J. Biol. Chem. *269*, 28506-28513.

Ahmed, M., Lyass, L., Markham, P.N., Taylor, S.S., Vazquez-Laslop, N., and Neyfakh, A.A. (1995). Two highly similar multidrug transporters of *Bacillus subtilis* whose expression is differentially regulated. J. Bacteriol. *177*, 3904-3910.

Allen, T.E. and Ullman, B. (1993). Cloning and expression of the hypoxanthine-guanine phosphoribosyltransferase gene from *Trypanosome brucei.* Nucleic Acids Res. *21*, 5431-5438.

Amabile-Cuevas, C.F. and Demple, B. (1991). Molecular characterization of the soxRS genes of *Escherichia coli*: two genes control a superoxide stress regulon. Nucleic Acids Res. *19*, 4479-4484.

Ansari, A.Z., Chael, M.L., and O'Halloran, T.V. (1992). Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. Nature *355*, 87-89.

Ansari, A.Z., Bradner, J.E., and O'Halloran, T.V. (1995). DNA-bend modulation in a repressor-to-activator switching mechanism. Nature *374,* 371-375.

Bailly, C., Mollegaard, N.E., Nielsen, P.E., and Waring, M.J. (1995). The influence of the 2-amino group of guanine on DNA conformation. Uranyl and DNase I probing of inosine/diaminopurine substituted DNA. EMBO J. *14*, 2121-2131.

Bailly, C., Payet, D., Travers, A.A., and Waring, M.J. (1996). PCR-based development of DNA substrates containing modified bases: an efficient system for investigating the role of the exocyclic groups in chemical and structural recognition by minor groove binding drugs and proteins. Proc. Natl. Acad. Sci. U S A 93, 13623-13628.

Baranova, N.N., Danchin, A., and Neyfakh, A.A. (1999). Mta, a global MerR-type regulator of the *Bacillus subtilis* multidrug-efflux transporters. Mol. Microbiol. 31,1549-1559.

Bjorkman, A.J. and Mowbray, S.L. (1998). Multiple open forms of ribose-binding protein trace the path of its conformational change. J. Mol. Biol. *279*, 651-664.

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

Blundell, T.L. and Johnson, L.N., eds. (1976). Protein Crystallography, Academic Press, New York.

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*, 4239-4245.

Bragg, W.L. (1913). The diffraction of short electromagnetic waves by a crystal. Proc. Camb. Phil. Soc. *17*, 43-57.

Branden, C. and Tooze, J. (1999). Introduction to protein structure. 16-17. Garland Publishing, Inc., NY.

Bruggemann, E.P., Currier, S.J., Gottesman, M.M., and Pastan, I. (1992). Characterization of the azidopine and vinblastine binding site of P-glycoprotein. J. Biol. Chem. 267, 21020-21026.

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

Brünger A.T. (1997). Free R value: cross-validation in crystallography. Meth. Enzymol. *277B*, 366-396.

Calladine, C.R. and Drew, H.R. (1992). Understanding DNA, Academic Press, New York.

Chen, J., Pongor, S., and Simoncsits, A. (1997). Recognition of DNA by single-chain derivatives of the phage 434 repressor: high affinity binding depends on both the contacted and non-contacted base pairs. Nucleic Acids Res. *25*, 2047-2054.

Chen, X., Tordova, M., Gilliland, G.L., Wang, L., Li Y., Yan, H., and Ji, X. (1998). Crystal structure of apo-cellular retinoic acid-binding protein type II (R111M) suggests a mechanism of ligand entry. J. Mol. Biol. *278*, 641-653.

Choi, K.Y. and Zalkin, H. (1990). Regulation of *Escherichia coli* pyrC by the purine regulon repressor protein. J. Bacteriol. *172*, 3201-3207.

Choi, K.Y. and Zalkin, H. (1992). Structural characterization and corepressor binding of the *Escherichia coli* purine repressor. J. Bacteriol. *174*, 6207-6214.

Choi, K.Y. and Zalkin, H. (1994). Role of the purine repressor hinge sequence in repressor function. J. Bacteriol. *176*, 1767-1772.

Compton, L.A., Mathews, C.K., and Johnson, W.C. Jr. (1987). The conformation of T4 bacteriophage dihydrofolate reductase from circular dichroism. J. Biol. Chem. *262*, 13039-13043.

Constantine, K.L., Friedrichs, M.S., Wittekind, M., Jamil, H., Chu, C.H., Parker, R.A., Goldfarb, V., Mueller, L., and Farmer, B.T. 2nd. (1998). Backbone and side chain dynamics of uncomplexed human adipocyte and muscle fatty acid-binding proteins. Biochemistry *37*, 7965-7980.

Cornwell, M.M., Safa, A.R., Felsted, R.L., Gottesman, M.M., and Pastan, I. (1986). Membrane vesicles from multidrug-resistant human cancer cells contain a specific 150- to 170-kDa protein detected by photoaffinity labeling. Proc. Natl. Acad. Sci. U S A 83, 3847-3850.

Creighton, T.E. (1993). Proteins, second edition. pp.334-336. W.H. Freeman and Company. New York.

Dickerson, R.E., Goodsell, D.S., and Neidle, S. (1994). "...the tyranny of the lattice...". Proc. Natl. Acad. Sci. USA *91*, 3579-3583.

Donaldson, L.W., Petersen, J.M., Graves, B.J., and McIntosh, L.P. (1996). Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. EMBO J. *15*, 125-134.

Drenth, J. (1994). Principles of protein x-ray crystallography, Springer-Verlag, New York.

Edgar, R. and Bibi, E. (1997). MdfA, an *Escherichia coli* multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. J. Bacteriol. *179*, 2274-2280.

Edgar, R. and Bibi, E. (1999). A single membrane-embedded negative charge is critical for recognizing positively charged drugs by the Escherichia coli multidrug resistance protein MdfA. EMBO J. *18*, 822-832.

Engh, R.A. and Huber, R. (1991). Accurate bond and angle parameters for x-ray protein structure determination. Acta Crystallogr. *A47*, 392-400.

Eriksson, A.E., Baase, W.A., Zhang, X.J., Heinz, D.W., Blaber, M., Baldwin, E.P., and Matthews, B.W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science *255*, 178-183.

Ewald, P.P. (1912). Das "reziproke Gitter" in der Structurtheorie. Zeitschrift für Kristallographie und Mineralogie, *56*, 129-156.

Ferré-D'Amaré, A.R. and Burley, S.K. (1994). Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. Structure *2*, 357-359.

Ferré-D'Amaré, A.R., Pognonec, P., Roeder, R.G., and Burley, S.K. (1994). Structure and function of the b/HLH/Z domain of USF. EMBO J. *13*, 180-189.

Fitzgerald, P.M.D. (1988). MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement. J. Appl. Cryst. *26*, 283-291.

Flomer, W.A., Kolis, J.W., and Pennington, W.T. (1996). Tris (tetraphenylphosphonium) decatelluriumtantallate dimethylformamide (1/1), a tetraphenylphosphonium salt of $TaTe_{10}^{3-}$. Acta Crystallogr. *C52*, 2445-2448.

Flower, D.R. (1996). The lipocalin protein family: structure and function. Biochem. J. *318*, 1-14.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. Science *270*, 397-403.

Furey, W., and Swaminathan, S. (1997). PHASES-95: a program package for the processing and analysis of diffraction data from macromolecules. Meth. Enzymol. *277B*, 590-620.

Gaudu, P., and Weiss, B. (1996). SoxR, a [2Fe-2S] transcription factor, is active only in its oxidized form. Proc. Natl. Acad. Sci. USA *93*, 10094-10098.

Germann UA. (1996). P-glycoprotein--a mediator of multidrug resistance in tumour cells. Eur. J. Cancer *32A*, 927-944.

Glasfeld, A., Schumacher, M.A., Choi, K.Y., Zalkin, H., and Brennan, R. G. (1996). A positively charged residue bound in the minor groove does not alter the bending of a DNA duplex. J. Amer. Chem. Soc. *118*, 13073-13074.

Goffeau, A., Park, J., Paulsen, I.T., Jonniaux, J.L., Dinh, T., Mordant, P., and Saier, M.H. Jr. (1997). Multidrug-resistant transport proteins in yeast: complete inventory and phylogenetic characterization of yeast open reading frames with the major facilitator superfamily. Yeast *13*, 43-54.

Gottesman, M.M., and Pastan, I. (1988). The multidrug transporter, a double-edged sword. J. Biol. Chem. *263*, 12163-12166.

Gottesman, M.M., and Pastan, I. (1993). Biochemistry of multidrug resistance mediated by the multidrug transporter. Annu. Rev. Biochem. *62,* 385-427.

Gralla, J.D. and Collado-Vides, J. (1996). Organisation and function of transcription regulatory elements. In *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. Neidhardt, F.C. et al (Eds). Washington, D.C. pp. 1232-1245.

Greenberger, L.M., Lisanti, C.J., Silva, J.T., Horwitz, S.B. (1991). Domain mapping of the photoaffinity drug-binding sites in P-glycoprotein encoded by mouse *mdr1b*. J. Biol. Chem. *266*, 20744-20751.

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*, 18665-18673.

Gros, P., Talbot, F., Tang-Wai, D., Bibi, E., and Kaback, H.R. (1992). Lipophilic cations: a group of model substrates for the multidrug-resistance transporter. Biochemistry *31*, 1992-1998.

Gunasekaran, K., Ramakrishnan, C., and Balaram, P. (1996). Disallowed Ramachandran conformations of amino acid residues in protein structures. J. Mol. Biol., *264*, 191-198.

Hahn, T. (Ed.) International Tables for Crystallography, D. Reidel Publishing Company, Dordrecht, Holland, 1983.

Hanna, M., Brault, M., Kwan, T., Kast, C., and Gros, P. (1996). Mutagenesis of transmembrane domain 11 of P-glycoprotein by alanine scanning. Biochemistry *35*, 3625-3635.

Harker, D. (1956). The determination of phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. Acta Crystallogr. *9*, 1-2.

Harris, K.D.M., Johnston, R.L., and Kariuki, B.M. (1998). The genetic algorithm: foundations and applications in structure solution from powder diffraction data. Acta Crystallogr. *A54*, 632-645.

He, B., Shiau, A., Choi, K.Y., Zalkin, H., and Smith, J.M. (1990). Genes of the *Escherichia coli pur* regulon are negatively controlled by a repressor-operator interaction. J. Bact. *172*, 4555-4562.

He B. Shiau A. Choi KY. Zalkin H. Smith JM., He B. Choi, K.Y., and Zalkin H. (1993). Regulation of *Escherichia coli glnB, prsA,* and *speA* by the purine repressor. J. Bact. *175*, 3598-3606.

Hendrickson, W.A. and Lattman, E.E. (1970). Representation of phase probability distributions for simplified combination of independent phase information. Acta Crystallogr. *B26*, 136-143.

Hidalgo, E., Ding, H., and Demple, B. (1997). Redox signal transduction via iron-sulfur clusters in the SoxR transcription activator. TiBS *22*, 207-210.

Hidalgo, E. and Demple, B. (1997). 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*, 1056-1065.

Higgins, C.F. and Gottesman, M.M. (1992). Is the multidrug transporter a flippase? TiBS *17*, 18-21.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123-138.

Holmes, D.J., Caso, J.L., and Thompson, C.J. (1993). Autogenous transcriptional activation of a thiostrepton-induced gene in *Streptomyces lividans*. EMBO J. *12*, 3183-3191.

Homolya, L., Hollo, Z., Germann, U.A., Pastan. I., Gottesman, M.M., and Sarkadi, B. (1993). Fluorescent cellular indicators are extruded by the multidrug resistance protein. J. Biol. Chem. *268*, 21493-21496.

Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. Science *268*, 1144-1149.

Hyde, S.C., Emsley, P., Hartshorn, M.J., Mimmack, M.M., Gileadi, U., Pearce, S.R., Gallagher, M.P., Gill, D.R., Hubbard, R.E. and Higgins CF. (1990). Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. Nature *346*, 362-365.

Ito, T., Yano, I., Tanaka, K., Inui, K.I. (1997). Transport of quinolone antibacterial drugs by human P-glycoprotein expressed in a kidney epithelial cell line, LLC-PK1. J. Pharmacol. Exp. Ther. *282*, 955-960.

Jackson, S.E., Moracci, M., elMasry, N., Johnson, C.M., and Fersht, A.R. (1993). Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. Biochemistry *42*, 11259-11269.

Jancarik, K. and Kim, S. (1991). Sparse Matrix sampling: a screening method for the crystallization of proteins. J. Appl. Cryst. *24*, 409-411.

Jen-Jacobson, L. (1995). Structural-perturbation approaches to thermodynamics of site-specific protein-DNA interactions. Methods Enzymol. *259*, 305-344.

Jones, T.Z., Zou, J.-Y., Cowan, S.W., and Kieldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110-119.

Juliano, R.L. and Ling, V. (1976). A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim. Biophys/ Acta *455*, 152-162.

Kim, Y.C., Grable, J.C., Love, R., Greene, P.J., and Rosenberg, J.M. (1990). Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. Science *249*, 1307-1309.

Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993a). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520-527.

Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993b). Crystal structure of a yeast TBP/TATA-box complex. Nature *365*, 512-520.

Kissinger, C.R., Gehlhaar, D.K., and Fogel, D.B. (1999). Rapid automated molecular replacement by evolutionary search. Acta Crystallogr. *D55*, 484-491.

Klyachko, K.A., Schuldiner, S., and Neyfakh, A.A. (1997). Mutations affecting substrate specificity of the *Bacillus subtilis* multidrug transporter Bmr. J. Bacteriol. *179*, 2189-2193.

Koshland, D.E.Jr. Catalysis in life and in the test tube. In Horizons in Biochemistry. Eds. M.Kasha and B. Pullman. Academic Press. New York, 1962.

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. *24*, 946-950.

Kussie, P.H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J., and Pavletich, N.P. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science *274*, 948-953.

LaLonde, J.M., Bernlohr, D.A., and Banaszak, L.J. (1994). The up-and-down beta-barrel proteins. FASEB J. *8*, 1240-1247.

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

LeMaster, D.M., Richards, F.M. (1985). NMR sequential assignment of *Escherichia coli* thioredoxin utilizing random fractional deuteriation. Biochemistry *24*, 7263-7268.

Lesser, D.R., Kurpiewski, M.R., and Jen-Jacobson, L. (1990). The energetic basis of specificity in the EcoRI endonuclease-DNA interaction. Science 250, 776-786.

Lewis, K. (1994). Multidrug resistance pumps in bacteria: variations on a theme. TiBS *19*, 119-123.

Li, A.J. and Nussinov, R. (1998). A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. Proteins *32*, 111-127.

Liu, R. and Sharom, F.J. (1996). Site-directed fluorescence labeling of P-glycoprotein on cysteine residues in the nucleotide binding domains. Biochemistry *35*, 11865-11873.

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

Loo, T.W. and Clarke, D.M. (1996). Inhibition of oxidative cross-linking between engineered cysteine residues at positions 332 in predicted transmembrane segments (TM) 6 and 975 in predicted TM12 of human P-glycoprotein by drug substrates. J. Biol. Chem. *271*, 27482-27487.

Lu, F., Schumacher, M.A., Arvidson, D.N., Haldimann, A., Wanner, B.L., Zalkin, H., and Brennan, R.G. (1998). Structure-based redesign of corepressor specificity of the

Escherichia coli purine repressor by substitution of residue 190. Biochemistry *37*, 971-982.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1998). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. Nature *352*, 497-505.

Lund, P. and Brown, N. (1989). Up-promoter mutations in the positively-regulated mer promoter of Tn501. Nucleic Acids Res. *17*, 5517-5527.

Lundblad, J.R., Laurance, M., and Goodman, R.H. Fluorescence polarization analysis of protein-DNA and protein-protein interactions. Mol. Endocrinol. *10*, 607-612.

Ma, D., Alberti, C., Lynch, H., 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*, 101-112.

Mandel-Gutfreund, Y., Margalit, H., Jernigan, R.L., and Zhurkin, V.B. (1998). A role for CH...O interactions in protein-DNA recognition. J. Mol. Biol. 277, 1129-1140.

Mandel-Gutfreund, Y., Schueler, O., and Margalit, H. (1995). Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. J. Mol. Biol. *253*, 370-382.

Marger, M.D. and Saier, M.H. Jr. (1993). A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. TiBS *18*, 13-20.

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,* 473-475.

Markham, P.N., LoGuidice, J., Neyfakh, A.A. (1997). Broad ligand specificity of the transcriptional regulator of the *Bacillus subtilis* multidrug transporter Bmr. Biochem. Biophys. Res. Commun. *239*, 269-272.

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

McPherson, A. (1985a). Crystallization of macromolecules: general principles. Meth. Enzymol. *114*, 112-119.

McPherson, A. (1985b). Use of polyethylene glycol in the crystallization of macromolecules. Meth. Enzymol. *114*, 120-124.

McPherson, A. (1985c). Crystallization of proteins by variation of pH or temperature. Meth. Enzymol. *114*, 125-127.

Meng, L.M., Kilstrup, M., and Nygaard, P. Autoregulation of PurR repressor synthesis and involvement of purR in the regulation of purB, purC, purL, purMN and guaBA expression in *Escherichia coli*. (1990). Eur. J. Biochem. *18*, 1873-1879.

Milla, M.E., and Sauer, R.T. (1995). Critical side-chain interactions at a subunit interface in the Arc repressor dimer. Biochemistry *34*, 3344-3351.

Morris, D.I., Speicher, L.A., Ruoho, A.E., Tew, K.D., Seamon, K.B. (1991). Interaction of forskolin with the P-glycoprotein multidrug transporter. Biochemistry *30*, 8371-8379.

Mowbray, S.L. and Cole, L.B. (1992). 1.7 Å X-ray structure of the periplasmic ribose receptor from *Escherichia coli*. J. Mol. Biol. *225*, 155-175.

Nagadoi, A., Morikawa, S., Nakamura, H., Enari, M., Kobayashi, K., Yamamoto, H., Sampei, G., Mizobuchi, K., Schumacher, M.A., Brennan, R.G., et al. (1995). Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain. Structure *15*, 1217-1224.

Navaza, J. (1994). AmoRe: an automated package for molecular replacement. Acta Crystallogr. *50*, 157-163.

Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I., and Aggarwal, A.K. (1995). Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. Science *269*, 656-663.

Neyfakh, A.A., Bidnenko, V.E., and Chen, L.B. (1991). Efflux-mediated multidrug resistance in *Bacillus subtilis*: similarities and dissimilarities with the mammalian system. Proc. Natl. Acad. Sci. USA *88*, 4781-4785.

Neyfakh, A.A. (1997). Natural functions of bacterial multidrug transporters. Trends Microbiol. 5, 309-313.

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

Oh, B.H., Pandit, J., Kang, C.H., Nikaido, K., Gokcen, S., Ames, G.F., and Kim, S.H. (1993). Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. J. Biol. Chem. *268*, 11348-11355.

Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M., and Zhurkin, V.B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc. Natl. Acad. Sci. U S A *95*, 11163-11168.

O'Neil, K.T., Shuman, J.D., Ampe, C., and DeGrado, W.F. DNA-induced increase in the alpha-helical content of C/EBP and GCN4. Biochemistry *30*, 9030-9034.

Otwinowski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachimiak. A., Marmorstein, R.Q., Luisi, B.F., and Sigler, P.B. (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution. Nature *335*, 321-329.

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

Paulsen, I.T., Skurray, R.A., Tam, R., Saier, M.H. Jr., Turner, R.J., Weiner, J.H., Goldberg, E.B., and Grinius, L.L. (1996b). The SMR family: a novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs. Mol. Microbiol. *19*, 1167-1175.

Paulsen, I.T., Brown, M.H., Littlejohn, T.G., Mitchell, B.A., and Skurray, R.A. (1996c). Multidrug resistance proteins QacA and QacB from Staphylococcus aureus: membrane topology and identification of residues involved in substrate specificity. Proc. Natl. Acad. Sci. U S A *93*, 3630-3635.

Parkhill, J. and Brown, N.L. (1990). Site-specific insertion and deletion mutants in the mer promoter-operator region of Tn501; the nineteen base-pair spacer is essential for normal induction of the promoter by MerR. Nucleic Acids Res. *18*, 5157-5162.

Patterson, A.L. (1934). A Fourier series method for the determination of the components on interatomic distances in crystals. Phys. Rev. *46*, 372-376.

Petersen, J.M., Skalicky, J.J., Donaldson, L.W., McIntosh, L.P., Alber, T., and Graves, B.J. (1995). Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. Science *269*, 1866-1869.

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*, 3975-3980.

Rhodes, D. and Burley, S.K. (1997). Protein-nucleic acid interactions. Curr. Opin. Struct. Biol. *7*, 73-75

Rhodes, G. (1993). Crystallography made crystal clear. Academic Press, New York.

Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H.A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell *87*, 1295-1306.

Riordan, J.R. and Ling, V. (1979). Purification of P-glycoprotein from plasma membrane vesicles of Chinese hamster ovary cell mutants with reduced colchicine permeability. J. Biol. Chem. *254*, 12701-12705.

Robinson, H., Gao, Y.G., Bauer, C., Roberts, C., Switzer, C., and Wang, A.H. (1998). 2'-Deoxyisoguanosine adopts more than one tautomer to form base-pairs with thymidine observed by high-resolution crystal structure analysis. Biochemistry *37*, 10897-10905.

Roepe, P.D., Wei, L.Y., Cruz, J., and Carlson, D. (1993). Lower electrical membrane potential and altered pHi homeostasis in multidrug-resistant (MDR) cells: further characterization of a series of MDR cell lines expressing different levels of P-glycoprotein. Biochemistry *32*, 11042-11056.

Rolfes, R.J. and Zalkin, H. (1988). *Escherichia coli* gene *purR* encoding a repressor protein for purine nucleotide synthesis. Cloning, nucleotide sequence, and interaction with the *purF* operator. J.Biol.Chem. *263*, 19653-61.

Rolfes, R.J. and Zalkin, H. (1990a). Purification of the *Escherichia coli* purine regulon repressor and identification of corepressors. J. Bacteriol. *172*, 5637-5642.

Rolfes, R.J. and Zalkin, H. (1990b). Autoregulation of *Escherichia coli* purR requires two control sites downstream of the promoter. J.Bacteriol. *172*, 5758-5766.

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*, 10685-10694.

Russell, D.R. and Bennett, G.N. (1982). Construction and analysis of in vivo activity of *E. coli* promoter hybrids and promoter mutants that alter the -35 to -10 spacing. Gene *20*, 231-243.

Sacchettini, J.C., and Gordon, J.I. (1993). Rat intestinal fatty acid binding protein. A model system for analyzing the forces that can bind fatty acids to proteins. J. Biol. Chem. *268*, 18399-18402.

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. Nat. Acad. Sci. USA *88*, 637-641.

Saier, M.H. Jr., Tam, R., Reizer, A., and Reizer, J. (1994). Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. Mol. Microbiol. *11*, 841-847.

Schultz, S.C., Shields, G.C., and Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science *253*, 1001-1007.

Schumacher, M.A., Macdonald, J.R., Bjorkman, J., Mowbray, S.L., and Brennan, R.G. (1993). Structural analysis of the purine repressor, an *Escherichia coli* DNA-binding protein. J. Biol. Chem. *268*, 12282-12288.

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*, 763-770.

Schumacher, M.A., Choi, K.Y., Lu, F., Zalkin, H., and Brennan, RG. (1995). Mechanism of corepressor-mediated specific DNA binding by the purine repressor. Cell *83*, 147-155.

Schumacher, M.A., Glasfeld, A., Zalkin, H., Brennan, R.G. (1997). The X-ray structure of the PurR-guanine-*purF* operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity. J. Biol. Chem. *272*, 22648-22653.

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*, 804-808.

Seoane, A.S. and Levy, S.B. (1995). Characterization of MarR, the repressor of the multiple antibiotic resistance (*mar*) operon in *Escherichia coli*. J. Bacteriol. *177*, 3414-3419.

Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A., and Sigler, P.B. (1994). Determinants of repressor/operator recognition from the structure of the *trp* operator binding site. Nature *368*, 469-473.

Shapiro, A.B., Corder, A.B., and Ling, V. (1997). P-glycoprotein-mediated Hoechst 33342 transport out of the lipid bilayer. Eur. J. Biochem. *250*, 115-121.

Sharom, F.J. (1997). The P-glycoprotein efflux pump: how does it transport drugs? J. Membr.Biol. *160*, 161-175.

Shen, D.W., Fojo, A., Chin, J.E., Roninson, I.B., Richert, N., Pastan, I., Gottesman, M.M. (1986). Human multidrug-resistant cell lines: increased *mdr1* expression can precede gene amplification. Science *232*, 643-645.

Sonveaux, N., Shapiro, A.B., Goormaghtigh, E., Ling, V., Ruysschaert, J.M. (1996). Secondary and tertiary structure changes of reconstituted P-glycoprotein. A Fourier transform attenuated total reflection infrared spectroscopy analysis. J. Biol. Chem. *271*, 24617-24624.

Spolar, R.S. and Record, M.T. Jr. (1994). Coupling of local folding to site-specific binding of proteins to DNA. Science *263*, 777-784.

Spronk, C.A.E.M., Slijper, M., van Boom, J.H., Kaptein, R., and Boelens, R. (1996). Formation of the hinge helix in the lac repressor is induced upon binding to the *lac* operator. Nat. Struct. Biol. *11*, 916-919.

Spurlino, J.C., Lu, G.Y., and Quiocho, F.A. (1991). The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J. Biol. Chem. *266*, 5202-5219.

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

Summers, A.O. (1992). Untwist and shout: a heavy metal-responsive transcriptional regulator. J. Bacteriol. *174*, 3097-3101.

Tang-Wai, D.F., Brossi, A., Arnold, L.D., and Gros, P. (1993). The nitrogen of the acetamido group of colchicine modulates P-glycoprotein-mediated multidrug resistance. Biochemistry *32*, 6470-6476.

Trakhanov, S. and Quiocho, F.A. (1995). Influence of divalent cations in protein crystallisation. Protein Science *4*, 1914-1919.

Tronrud, D.E., TenEyck, L.F., and Matthews, B.W. (1987). An efficient general-purpose least squares refinement program for macromolecular structures. Acta Crystallogr. *A43*, 489-501.

Tronrud, D.E. (1997). TNT refinement package. Meth. Enzymol. *277B*, 306-319.

Uesugi, M., Nyanguile, O., Lu, H., Levine, A.J., and Verdine, G.L. (1997). Induced alpha helix in the VP16 activation domain upon binding to a human TAF. Science *277*, 1310-1313.

Ullman B. (1995). Multidrug resistance and P-glycoproteins in parasitic protozoa. J Bioenerg Biomembr. *27*, 77-84.

Utschig, L.M., Bryson, J.W., and O'Halloran, T.V. (1995). Mercury-199 NMR of the metal receptor site in MerR and its protein-DNA complex. Science *268*, 380-385.

Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Methods in Enzymology *115*, 90-112.

Weickert, M.J. and Adhya, S. (1992). A family of bacterial regulators homologous to Gal and Lac repressors. J. Biol. Chem. *267*, 15869-15874.

Wilson, H.R. and Turnbough, C.L. Jr (1990). Role of the purine repressor in the regulation of pyrimidine gene expression in *Escherichia coli* K-12. J.Bacteriol. *172*, 3208-3213.

Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K., and Wilson, K.S. (1993). The crystal structure of

EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. EMBO J. *12*, 1781-1795.

Xuong, N.H., Nielsen, C., Hamlin, R., and Anderson, D.J. (1985). Strategy for data collection from protein crystals using a multiwire counter area detector diffractometer. J. Appl. Cryst. *18,* 342-350.

Zalkin, H. and Nygaard, P. Biosynthesis of purine nucleotides. In: *Escherichia coli* and Salmonella typhimirium: Cellular and Molecular Biology, Second Ed. American Society for Microbiology, Washington, D.C., pp.561-579, 1996.

Zalkin, H. and Dixon, J.E. *De Novo* Purine Nucleotide Biosynthesis. In: Progress in Nucleic Acid Research and Molecular Biology, Vol. 42. Academic Press, NY pp.259-287, 1992.

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 Science *6*, 2465-2468.