Analysis of Substitutions of Glutamate 90 and Glutamate 120 in the Multidrug Resistance Regulator QacR Reveal New Modes of Ligand Binding

Benjamin Ellis Brooks

A Dissertation

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

in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy
November 2005

School of Medicine
Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is certify that the Ph.D. thesis of

Benjamin Ellis Brooks

has been approved

[redacted]
Professor in charge of thesis

[redacted]
Member

[redacted]
Member

[redacted]
Member

[redacted]
Member

Table of Contents

Catalog of Figures

Chapter 1: Introduction to QacR and Multidrug Resistance

Chapter 2: Introduction to Macromolecular Crystallography

Chapter 3: Methods

Chapter 4: X-ray Crystallography Reveals New Binding Modes for Ethidium Bromide and Rhodamine-6G with a Glutamate to Glutamine Substituted QacR at Position 90

Chapter 5: X-ray Crystallography Reveals a New Binding Mode for Ethidium Bromide with a Glutamate to Glutamine Substituted QacR at Position 120

Chapter 6: X-ray Crystallography Suggests QacR Has Multiple Binding Positions for Diamidine Compounds DB75 and DB 359

Abbreviations

Bmr	*Bacillus subtilis* multidrug transporter
BmrR	repressor of the *Bacillus subtilis* multidrug transporter
BRC	C-terminal portion of BmrR
CprB	A-factor receptor homologue from *Streptomyces coelicolor*
d_{min}	smallest lattice plane distance, or resolution
DNAse I	deoxyribonuclease I
Dq	dequalinium or 1,1'-Decamethylenebis(4-aminoquinaldinium)
Et	ethidium or 3,8-Diamino-5-ethyl-6-phenylphenanthridinium
EtBr	ethidium bromide
EthA	a monooxygenase that activates the prodrug ethionamide
EthR	repressor of EthA expression
HTH	helix-turn-helix
ITC	isothermal titration calorimetry
MAD	multi-wavelength anomalous diffraction
MarA	activator of multidrug resistance transporter complex AcrAB-TolC
MarR	repressor of MarA expression
MdfA	*E. coli* multidrug transporter
MerR	regulator of the mercury resistance operon
MFS	major facilitator superfamily
MG	malachite green or *N,N,N',N'*-Tetramethyl-4,4'-diaminotriphenylcarbenium
MIR	multiple isomorphous replacement
MR	molecular replacement

MRSA	methicillin resistant *Staphylococcus aureus*
PBP	penicillin binding protein
PXR	pregnane X receptor
PXR-LBD	PXR ligand binding domain
QAC	quaternary ammonium compound
QacA/B	quaternary ammonium compound transporter (A and B are isoforms)
QacR	repressor of the expression of the ammonium compound transporter
R6G	rhodamine-6G or xanthylium, 9-(2-(ethoxycarbonyl)phenyl)-3,6-bis(ethylamino)-2,7-dimethyl
Sig, σ, and Σ	sigma, or error
TetA(K)	tetracycline efflux pump
TetR	tetracycline efflux pump repressor
TM	transmembrane
TMD	transmembrane domain
TPP	tetraphenyl phosphonium

Acknowledgements

My work would not be possible with out the efforts of many other people. My dissertation would not be possible with out the help of Katie Hardie from the laboratory of Ronald Skurray. She made all the constructs and did the bulk of the affinity measurements for the ligands mentioned in the dissertation. I truly appreciate her efforts. Moreover, my research is built on other groundbreaking work done in Ronald Skurray's laboratory, particularly the work of Melissa Brown and Steve Grkovic. Similarly, I am indebted to Maria Schumacher who not only initiated the structural work on QacR with her studies of the ligand bound and DNA bound WT-QacR structures, she was also helpful throughout my graduate career especially with this project. Specifically she collected and processed the E90Q-QacR-R6G complex data that ended up being critical to this dissertation. She also identified K67 as a potentially important crystal contact. I would also like to acknowledge Jay Nix at the Molecular Biology Consortium beam line at the Advanced Light Source for taking the time to reprocess my data with which I was having difficulty. Other important contributors to the QacR project whose work is critical to this dissertation and our current understanding of QacR are Marshal Miller, David Murray, and Jason Schuman. Furthermore, the works of Ryan Watkins on PXR proved to be an important precedent for multidrug-binding proteins and gives further context for the results in this dissertation.

I am also indebted to Richard Brennan for taking me into his laboratory. He is a great mentor and I learned a great amount from him. He fostered an engaging scientific community in the lab as well as an engaging social environment. I will miss both. I also appreciate that he has been good-natured about my frequent intrusions into his office. He gave me an interesting project without which I would obviously not have written this dissertation. I am planning on pursuing similar projects in my post-graduate careers.

Similarly I appreciate all the lab members and friends over the years that have been incredibly helpful in my graduate career. Although I cannot think of someone who

was not helpful at some point in time, I would especially like to thank Kate Newberry, Joy Huffman, Dave Murray, Kate Hoffman, Caleb Bashor, Jason Schuman, Nicola Carter, and Sigrid Roberts for their knowledge and aid. The helpful friends that reviewed my thesis to make sure it was grammatically correct and understandable to someone other than myself, included Jennifer Hill, Kate Hoffman, and especially Rachel Dresbeck (from the Research Development Office).

The support staff in the department need mentioning as they are always helpful and are the unsung hero and heroines that get things done around the department.

I would also like to mention my other mentors at OHSU who have helped me shape my approach to science. My first mentor Dave Farrens gave me my first science job following college. I was and still am inspired by the rigor and creativity of the research pursued in his laboratory. Buddy Ullman, one of my first graduate school mentors, and his laboratory have were also incredibly encouraging and I appreciate among other things the molecular biological techniques I learned from my experiences in his laboratory. Furthermore, I appreciate the help and guidance of the members of my research advisory committee: Richard Brennan, Maria Schumacher, David Farrens, and James Lundblad. I would also like to note that I appreciate that Arthur Glasfeld along with my research advisory committee members for taking the time to be on my examination committee with a special thanks to Dr. Lundblad for chairing it.

My family probably does not realize how important they were for this process. I hope I make them proud. From starting my early interest in science by doing a report on the nitrogen cycle with my dad in middle school to my mom's constant encouragement I appreciate them both deeply. My fiancé, Jen, has been my keel to keep me even through the discouraging times, but also gives me room to make my own mistakes. I am not sure what I would do without her.

I appreciate everyone who has had the patience and interest in encouraging me and helping me to get here. Thank you.

Abstract

The repressor QacR regulates the expression of the multidrug exporter QacA, which mediates antiseptic resistance of *Staphylococcus aureus* to a wide variety of molecules. The ligands of both of QacR and QacR are lipophilic cations. In crystal structures of wild-type-QacR-ligand complexes, glutamates 90 and 120 appear to be critical for binding rhodamine-6G, ethidium, malachite green, and dequalinium (*1*). In order to determine if the negative charges of these glutamates plays a role in ligand binding, these residues were substituted with glutamines. The interactions of these substituted proteins with the structurally characterized QacR ligands: ethidium, malachite green, dequalinium, and rhodamine-6G were studied. The affinities of the substituted QacR to each ligand were measured with tryptophan fluorescence quenching, fluorescence polarization, and isothermal titration calorimetry. The structures of the protein ligand complexes were studied with X-ray crystallography

The structures of the E90Q-QacR-R6G complex and E120Q-QacR-Et complex indicated two binding positions were available to each of the ligands in the substituted proteins compared to the one binding position observed in the WT-QacR. As the binding affinity of the substituted proteins for these ligands did not dramatically change from that of the WT, these new binding positions must be accessible in the WT-QacR. Furthermore, the structures of the WT-QacR with DB75 and DB359 indicate that there may be many equivalent binding positions for these ligands as well. This research further demonstrates how the QacR binding pocket binds a multitude of disparately shaped ligands. Not only does this research confirm that the binding pocket changes to accommodate various ligands, but it illustrates that the same ligand can interact with the protein in more than one way.

Chapter 1: Introduction to QacR and Multidrug Resistance

Drug resistance in clinically relevant bacteria has been mounting since the advent of antimicrobials. It is now realized that some strains have become resistant to many drugs and antiseptics used clinically, and these are considered multidrug resistant (MDR). Approximately 40-60% of the *Staphylococcus aureus* strains found in the United States and the United Kingdom are methicillin resistant (MRSA) and usually multidrug resistant (*2; 3*). Slowly MRSA strains are becoming resistant to vancomycin, the drug of last resort (*4*). MRSA are also growing resistant to antiseptics (*5*).

This antiseptic resistance in *S. aureus* appears to be primarily due to the expression of the plasmid encoded drug resistance protein, quaternary-ammonium-compound transporter (QacA). Harnessing an internally directed proton gradient, QacA transports quaternary amine compounds (QACs) and a wide variety of lipophilic cations, including known hospital disinfectants, out of *S. aureus*. Hydrophobicity and charge are the only two understood similarities of its substrates. The substrate profile of this transporter is almost identical to that of QacR, the protein that regulates the expression of this transporter (*6*).

The transporter and its regulator may also share similarities in how they recognize the same ligands. Therefore, we may be able to gain insight to the transporter by studying its regulator. This is of particular interest as structural studies on QacA have proceeded slowly as it is a membrane bound protein, which makes it difficult to crystallize for X-ray crystallographic analysis. One of other reasons we are interested in studying ligand interactions with QacR is to understand the differences between ligand

recognition by the transporter and its regulator. Such knowledge may help us design ligands that induce expression of the transporter without being themselves transported, thereby straining the proton gradient that is critical to the survival of the bacteria.

Our main interest with QacR is purely to gain understanding on ligand binding with a multidrug binding protein. On one hand, we would like to understand how do these proteins recognize so many various molecules. On the other hand, we would like to understand how the ligand-binding pocket determines the limited ligand selectivity of QacR.

Previous studies on QacR by Dr. Maria Schumacher and David Murray indicated that five glutamates appeared to be important for ligand binding to QacR (Figures 1 and 2) (*7; 8*). The ligands for QacR and QacA are various monovalent and bivalent, disparately shaped molecules including ethidium (Et), rhodamine-6G (R6G), malachite green (MG), dequalinium (Dq), and pentamidine, all of which bound close to glutamates 57, 58, 63, 90, and 120. Otherwise, these ligands were found to bind in distinct locations and geometries in the binding pocket.

In order to understand the roles of the binding-pocket glutamates of QacR in ligand binding, we substituted each with a glutamine and studied the resultant substituted proteins. Katie M. Hardie in Ronald Skurray's laboratory at the University of Sydney provided the plasmids coding the substituted QacRs and performed the fluorescence quenching assays with all the substituted proteins. As part of the Brennan group, Dr. Jason Schuman and I performed ITC binding experiments for Dq. The study of the interactions of the substituted proteins was divided between Dr. Schuman and I. Dr. Schuman studied the structures of the E57Q and E58Q substituted QacRs with the X-ray

crystallography; I studied the structures of ligands bound to the E90Q and E120Q substituted QacRs. Because the novel-binding mode of pentamidine suggested that there were new possible binding modes available to QacR that included new interactions with amino acids, I also studied the binding of new types of ligands, experimental DNA binding drugs DB75 and DB359, to the unsubstituted QacR.

This dissertation begins with two introductory chapters that discuss multidrug resistance and X-ray crystallography, respectively. The experimental section of this thesis will be divided into four chapters. As many of the techniques used are identical from chapter to chapter, the third chapter describes the methods. Thus, many of the methods described in the later experimental chapters will be brief with the understanding that further detail may be obtained in this chapter. Following the methods chapter are two chapters discussing the role of glutamate 90 in ligand binding and the role of glutamate 120 in ligand. The final experimental chapter will discuss the disorder seen in the binding pocket of QacR complexed with DB75 or DB359 (Figure 3). The last chapter of the thesis is a discussion, which will draw together the ideas of the experimental chapters and discuss experiments that would further our understanding of QacR.

In short, we discovered that a single ligand can have multiple binding modes accessible for binding to QacR. The structures of the E90Q-QacR-R6G complex and E120Q-QacR-Et complex indicated two binding positions were available to each of the ligands in the substituted proteins compared to the one binding position observed in the WT-QacR. As the binding affinity of the substituted proteins for these ligands did not dramatically change from that of the WT these new binding positions must be accessible

in the WT-QacR. Furthermore, the structures of the WT-QacR with DB75 and DB359 indicate that there may be many equivalent binding positions for these ligands.

Mechanisms of Drug Resistance

Bacteria and other organisms use one or more of a number of different mechanisms to achieve drug resistance. The different methods of action for drug resistance can be split into two loose categories: action focused on the drug's target and action focused on the drug. The actions focused on the target for drug resistance may include altering the target to avoid drug binding or activity, duplicating the target, increasing repair of the target, or duplicating its function by another protein. The actions focused on the drug for drug resistance may include enzymatically altering the drug, reducing enzymatic activation of a prodrug, sequestering the drug, inhibiting its uptake, or increasing its export.

Methicillin resistance provided by the penicillin binding protein 2 (PBP2) in methicillin resistant *S. aureus* (MRSA) is illustrates a case of multidrug resistance through alteration of the target. The PBP2 is a critical component to cell wall biosynthesis in bacteria that is targeted through suicide inhibition by β-lactam drugs such as penicillin (Figure 4A). The β-lactams have a poor affinity for the PBP2 and depend on the reactivity of the enzyme to fix the drug to the active site. MRSA take advantage of this disparity by using a mutant form of PBP2 (PBP2a) that lowers the acylation activity of the binding site for substrates and inhibitors alike (*9*). This ensures fewer successful inactivation events by any β-lactam, thus allowing the enzyme to function, though at a reduced rate.

Another example of resistance to the same class of drugs illustrates how action on the drug itself can lead to drug resistance. To deactivate β-lactam based drugs, Bacteria developed specialized enzymes called β-lactamases that open the β-lactam ring of β-lactam-based drugs (Figure 4B). Without this β-lactam ring the drugs are deactivated and not able to act as a suicide inhibitors to the penicillin binding protein. Similarly, failure to activate a prodrug such as the nucleoside analogue gemcitabine provides drug resistance for leukemia cells. Gemcitabine requires phosphorlylation by deoxycytidine kinase in order to become an effective anticancer drug. Thus, drug resistant leukemia cell lines express deoxycytidine kinase to a lesser degree, thereby lowering the amount of gemcitabine activation (*10*).

A common element of these examples of drug resistance is that they are very specific to the drug, the target, or both. For most drug resistance mechanisms—such as enzymatic activation or inactivation; target modification, duplication, or repair—action on a drug is specific for the target and action on the target is specific to a set of similar drugs. However, drug efflux can provide bacterial resistance to a variety of antimicrobial molecules that have only a few structural qualities in common. For example, QacA exports a wide variety of lipophilic cations out of the bacteria. These molecules may have nothing in common other than the lack of a negative charge (*11; 12*). *Bacillus subtilis* multidrug transporter, Bmr, has a similar substrate profile but also transports a number of neutral and zwitterionic drugs (*13*). Furthermore, proteins, such as the *E. coli* multidrug transporter, MdfA, and the export an extremely wide variety of substrates even compared to other multidrug transporters. The substrates for MdfA include lipophilic, aromatic, non-aromatic, basic, zwitterionic, and neutral molecules such as a number of

antibiotics and disinfectants (*14*). As multidrug efflux can have such a wide range of substrates and is a common form of drug resistance from prokaryotic to eukaryotic life, the study of multidrug transporters and the proteins that regulate their expression is an important focus of scientific research.

Efflux

Multidrug efflux proteins are responsible for much of the multidrug resistance seen in *S. aureus*. The first multidrug efflux proteins reported for *S. aureus* include QacA and Smr (QacC), which have been found on several multidrug-resistance related plasmids (*5*). These proteins protect bacteria from disinfectants and antiseptics that are often found in hospitals. Of the 497 clinical isolates of *S. aureus* from 24 hospitals in Europe, 42% of these had QacA or the homologue QacB genes (*15*). Unfortunately, these antiseptic genes are often found in isolates of multidrug resistant MRSA. In the same study of European *S. aureus*, of the MRSA isolates 63% of these had QacA or B genes (*15*). In Asia 42% of 894 MRSA isolates carried genes for QacA or B genes (*5*). Similarly, the prototypical QacA carrying drug resistance marker also confers resistance to antibiotics, including gentamycin, tobramycin, and kanamycin (*16*). Thus, with this plasmid being passed between bacteria, drug resistance also selects for antiseptic resistance and antiseptic resistance selects for drug resistance.

QacA and B are completely homologous, except for one residue (Figure 5). Thus, both QacA and QacB are members of the 14 TMD major facilitator superfamily (MFS) and antiport drugs against the proton gradient that lies across the bacterial membrane (*11; 17*). In fact, analysis of the genes and the distribution of their plasmids suggest that QacA evolved from QacB (*18*). There are 7 pb different between QacA and QacB, but they

only code for one amino acid change at residue 323 (*11*). In QacA, this residue is an aspartate and in QacB, this residue is a glycine (*11*). This small change delineates the difference between the bivalent cation binding profile of QacA and the monovalent cation binding of QacB. This difference suggests that the negatively charged aspartate 323 is important for binding the second positive charge of a bivalent cation (*19*). However, neither QacA nor QacB transport trivalent cations, which could be important for future drug development.

Overall, QacA is able to transport over 30 cationic, lipophilic, antimicrobial compounds that belong to 11 different chemical classes. Both monovalent and bivalent cations are competitive, so they must be transported via the same mechanism (*17*). Furthermore, QacA uses both the pH gradient component and the charge gradient component of the proton gradient (*17*). Since QacA antiports positively charged molecules, the analysis of the proton gradient suggests that, for every monovalent cation transported out of the bacteria by QacA, at least 2 protons are transported into the bacteria. QacA also has been shown to remove specific ligands from the inner leaflet of the cytoplasmic membrane (*17*). The universality of this mechanism for all QacA ligands is unknown, but as all of the ligands are lipophilic and thus likely to be soluble in the membrane, this mechanism is an attractive hypothesis.

Regulation of Multidrug Resistance Transporters

With the exception of translational regulation of TetA(L) and TetA(K) and the unregulated members of the SMR family, the expression of each of the multidrug resistance proteins is regulated at the transcriptional level (*20; 21*). Expression may be modulated up or down through the actions of transcriptional repressors and activators.

These regulators generally belong to one of these four regulatory protein families: AraC, MarR, MerR and TetR. The assignment of these families is based solely on similarities in the DNA binding domains of the protein. All known MDR regulators possess the α-helix-turn-α-helix (HTH) DNA binding motif which may or may not be embedded into larger DNA binding domains (*20*). The HTH motif serves as an integral structural and functional motif that orients the second helix (the DNA sequence recognition helix) towards the major groove of the DNA.

AraC Family of Drug Resistance Regulators

The structures of the proteins MarA, Rob, and SoxS best define the first family of proteins, AraC. All three of these regulators are important activators of the expression of the MDR efflux protein complex, AcrAB/TolC (Figure 6). Overexpression of any of these activators has been shown to activate the transcription of components of AcrAB/TolC. These proteins bind as monomers to marboxes located near the promoters of the target genes. The marboxes are degenerate 20-bp elements lacking in symmetry and found scattered throughout the genome, as MarA is a global activator that regulates more than 60 genes.

The members of this family bind to DNA as a monomer that contains two HTH motifs. MarA inserts the two sets of HTH motifs in two adjacent portions of the major groove (Figure 7). The α-helix connecting the two domains is rigid thus causing the DNA to accommodate and bend by ~35° (*22*). There are a number of non-specific backbone contacts and several hydrogen bonds made by both recognition helices with several bases of the DNA (*23*). The two HTH motifs in the N-terminal domains of both MarA and Rob are superimposable. Remarkably, the crystal structure of Rob, with its

binding element *micF*, indicates that the C-terminal HTH of the two HTH domains do not insert into the major groove and make only one contact with the DNA, the DNA remaining, unbent (Figure 7)(*24*). However, the DNA used in the to cocrystallize with the MarA and Rob proteins were from different promoter classes, so it is possible that determination of whether the C-terminal HTH inserts into the major groove may be dependent on the DNA rather than the protein (*20*). Structurally, the biggest difference between the MarA family regulators is that MarA and SoxS have no apparent ligand-binding domain and Rob has a putative C-terminal binding domain (Figure 7) (*20*).

MarR Family of Drug Resistance Regulators

MarR regulates AcrAB/TolC gene expression by regulating the expression of MarA (Figure 6). Unlike MarA this protein contains a drug-binding domain that binds a broad range of ligands, including a number of known antibiotics. MarR is a homodimeric protein each subunit of which contains a DNA-binding/ligand binding domain and a dimerization domain (Figure 8A). The overall structure gives the appearance of an upward pointing triangle with the DNA/ligand-binding domains forming the flat bottom. Extending at an angle from the DNA-binding domains, the dimerization domains reach towards each other and cross each other in a nearly orthogonal manner. The α1, α5, and α6 form the dimerization domain of each subunit each of which interdigitates with the other to form well-packed hydrophobic core that buries a total of 3,570 $Å^2$ (*25*).

The DNA-binding domain consists of a winged-helix motif where helices 3 and 4 constitute the HTH motif surrounded by the two β-sheet wings (*20*). Modeling and mutagenesis indicate that the α4 from each subunit likely to interact with the major groove of the 21 base pair operator between the -10 and -35 regions of the *MarRAB*

promoter. However, the half sites for the operator are on opposite sides of the DNA helix. Proposed scenarios for DNA binding include the change of the conformation of the protein to position the DNA binding helices on either side of the DNA and the binding of two MarR dimers, one to either face of the DNA. The MarR inducer, salicylate, was shown to bind to proximal to α4 in the DNA binding domain in two sites per dimer (Figure 8B). Thus, inducer binding is likely to symmetrically affect DNA binding though direct interactions with the DNA binding helix α4 in each subunit.

MerR Family of Drug Resistance Regulators

The structures of the repressor for the multidrug transporter Bmr, BmrR, is a good representative of the regulator family, MerR. The binding domain of MerR family members belongs to the winged-helix superfamily, which includes a HTH (including the region α1 through α2) and two β-sheet DNA binding elements (Figure 9A). Members of this family also bind as a dimer to an imperfect inverted repeat located between the -10 and -35 hexamers of a target promoter that exhibits an unusually large spacing of 19 bp (*26*). The 19-bp spacing creates a promoter incompatible with RNA polymerase binding by placing the -10 and -35 regions of the promoter on opposite sides of the DNA helix. A deletion of two base pairs in the spacer regions of the prompters of MerR members has been shown to create constitutively active promoters by bringing the -10 and -35 regions closer together and on one side of the helix of the DNA (*27*). Similarly, when BmrR binds drugs, it activates the promoter by partially untwisting the promoter DNA, bringing the -10 and -35 regions closer together and on one side of the promoter (Figure 9B)(*28*). Thus drug binding to BmrR allows the expression of the efflux protein Bmr, which like the transporter QacA, exports lipophilic cations (Figure 10).

These lipophilic cationic ligands bind in a portion of BmrR that is hydrophobic but contains the buried charge E253. The structure of the ligand binding domain or C-terminal portion of BmrR (BRC) occupied by tetraphenyl phosphonium (TPP) illustrates a distance of only 3.2 Å between the closest ring and an oxygen atom from the carboxylic acid group at the end of E253 (*29*). Furthermore, a recent structure of the full-length BmrR-R6G complex indicates that the R6G basically occupies the same position as TPP and comes to a distance of 3.7 Å of E253 (Figure11)(*30*). Both of these structures indicate that the negative charge of the glutamate may be important for countering the buried charge of the ligand. Otherwise residues of the binding pocket are uncharged, as it is composed of aromatic and lipophilic ligands, including tyrosines, phenylalanines, isoleucines, and valines. Thus, in addition to the charge-charge interactions, the interactions between the drugs and the pocket include several van der Waals contacts with the various hydrophobic residues in the pocket and the rings of TPP and R6G stack against tyrosines and phenylalanines in the pocket.

Analyses of the ligand-free BRC and the BRC-TPP complex illustrate a possible mechanism of activation of the protein as a result of drug binding. Without drug in the binding pocket, Y152, the first residue of the α-helix following the loop covering the binding pocket, occupies this space (*29; 30*). Thus, in the ligand-free BRC structure, this tyrosine contributes one of the internal hydrogen bonds to the buried carboxylate of E253 and therefore occupies space that would otherwise be taken by ligand. All of the subsequent structures of BmrR with DNA include ligand and are in the transcription-activated conformation in which Y152 and its helix have moved 11.6 Å away from the positions in the apoBRC structure (Figure 12)(*30*). This indicates that drug binding

replaces Y152 in the binding pocket and the new conformation adopted by the protein confers transcriptional activation of the DNA. This new structure of the protein bunches up and twists the DNA so that the -10 and -35 elements of the promoter are at the proper distance and orientation for RNA polymerase to recognize them.

TetR Family of Multidrug Resistance Regulators

The final major family of transcription regulators for multidrug resistance transporters is the TetR family, the family of regulators to which QacR belongs. It is a large structural family of over 2,500 members (*31*). Membership to this family is defined by high sequence similarity in the DNA binding domains of these proteins, as the rest of the protein tends only to be structurally homologous (*31*). The specific functions of only about 85 of these are known. However, this family of proteins is known to regulate a wide range of cellular functions, including multidrug resistance, catabolism, antibiotic biosynthesis, osmotic stress response, and bacterial pathogenicity (*31*). More generally, they appear to be involved in the adaptation to complex and changing environmental conditions. For example, the TetR family member CprB from *Streptomyces coelicolor* A3(2) serves as a negative regulator for morphogenesis and secondary metabolism, such as antibiotic production (*32*).

Many of the other characterized TetR family members are involved in drug resistance, and, in one case, activation of a prodrug. EthR regulates the expression of the monooxygenase EthA, which is required to activate the anti-*Mycobacterium tuberculosis* prodrug ethionamide. Inducing EthR with other molecules in order to increase the activity of EthA increases the effectiveness of ethionamide (*33*). In order to protect

bacteria from drugs, QacR regulates of multidrug resistance in *S. aureus*, and TetR regulates protection from tetracycline-Mg^{2+} in Gram-negative bacteria.

TetR, the repressor of the tetracycline transporter, TetA, is the thoroughly characterized member of the family. TetA is a tetracycline-Mg^{2+}/H^{+} antiporter and high levels of expression of TetA have been shown to cause a lethal collapse of the cytoplasmic membrane potential (Figure 13)(*34*). To protect the bacteria from this possibility, TetR tightly represses the expression of TetA by binding to the 15bp *TetO* operator site with an association constant, K_a, on the order of 10^{11} M^{-1}. As tetracycline-Mg^{2+} is a potent inhibitor of ribosomal activity, with a K_d of approximately 1 μM, repression is sensitively regulated. At concentrations of 1 nM tetracycline-Mg^{2+}, TetR is induced, reducing its affinity for *TetO* by 9 orders of magnitude (*35-37*). Thus, the TetR-*TetO* system is the most efficiently inducible system for regulation of gene expression and therefore has gained the interest of many biologists as an inducible system for transcriptional regulation in eukaryotes (*35-37*). It is the interest in understanding its efficiency as a repressor, as well as its use as a biological tool that has inspired intense genetic and biochemical characterization of this system—even so far as to create a "reverse repressor" that binds more tightly to the operator site as tetracycline binds (*38*).

The repressor of QacA/B from *Staphylococcus aureus*, QacR, is the best structurally characterized repressor of this family. Including the structures in this thesis, there are at least twenty structures of QacR and various mutants, with a number of different ligands. The laboratories of Ronald Skurray and Richard Brennan are characterizing many mutants of this protein using genetic, biochemical, and structural

techniques, in an effort to understand the mechanics of QacR ligand binding and induction.

QacR regulation of gene expression compared to TetR is much less efficient. The induction of QacR allows for only 6-fold better expression of the multidrug transporter QacA, which is already expressed to a degree by inefficient repression (Figure 14)(*39*). This weak repression may be partly due to the location of the operator, which indicates that QacR does not appear to affect RNA polymerase binding but rather affects the processivity of polymerase. Another part of the weak repression appears to be due to the sequence of the 6 bp region between the 15 bp half sites of the unusually large QacR operator (Figure 15A). Repression by QacR can be tightened by a simple transversion of two bases in the spacer region (*40*). This indicates that the operator has evolved to be leaky, suggesting that continual low-level expression of this gene is more favorable than a more highly controlled system. It is still unknown exactly what the natural ligands for the QacA/R system are, but this leaky system should apply some innate protection from these molecules.

Determination of what differentiates the ligand profile of QacR and the ligand profile of QacA could be pharmacologically important. Thus far, of all the QacR ligands known, methyl-green is the only molecule that interacts with QacR but not QacA (*6*). Furthermore, QacR has measurable affinities for the trivalent molecule CGP 40215A (Schuman, J., unpublished). Though we do not know the affinity of QacA for CGP 40215A, QacA has only been shown to transport monovalent or bivalent cations thus far. These studies hint at the possibility that a drug may be created that interacts with QacR but not QacA such that QacA is overexpressed by QacR induction but cannot clear the

drug. The overabundance of QacA could then destroy the membrane potential and kill the cell, as was observed with the overexpression of TetA.

The structures of the TetR family members that have been solved thus far are composed entirely of α-helices. The proteins tend to range in length from 180 to 220 amino acids and contain from 9 to 10 α-helices (Figure 16 A&B). QacR is a homo dimer of two 188-residue polypeptides that each fold into nine α-helices (Figure 16A). Structural family member EthR also folds as a dimer with nine α-helices per subunit, while TetR and CprB both have ten α-helices (Figure 16A)(*32; 38*). Other than the extra helix, the topology difference between QacR and TetR consists of the switching of these last two helices in TetR. Helix α10 in TetR replaces the position of α9 in QacR. Interestingly, even with varying numbers of α-helices, both EthR and CprB have the same topology as QacR (*32; 33; 41*). Given what we know about the structures of the TetR family members thus far, QacR is more representative of the structural family than TetR.

The TetR members contain a DNA binding domain and a dimerization/ligand-binding domain. The DNA-binding domain, which is homologous in all of the TetR members, is defined by the first three helices (Figure 16 A&B). The actual HTH DNA binding motif is defined by the stretch of amino acids from α2 through α3, illustrating the canonical HTH DNA binding motif. Helices 4 and 4' subsequently connect these domains to the dimerization/ligand-binding domain. The following 5 or 6 C-terminal α-helices form the dimerization and drug-binding domain. In QacR, the terminal helices, α8 and α9, from each subunit make up the dimerization interface. In contrast, the dimer interface in TetR is composed primarily with α8 and α10.

There is no consensus for the mode of DNA binding by TetR members. A single TetR dimer binds to a single 15 bp inverted repeat. The crystal structures of the QacR-DNA complex illustrate that two QacR dimers bind to a pair of inverted repeats that span 28 bp (Figure 15B) (*42*). ITC and gel filtration chromatography confirm the stoichiometry and indicate that the binding is cooperative. Similarly, stoichiometry measurements of DNA binding to the TetR family member MtrR indicate that it binds as two dimers to DNA (*43*). Further analysis of the crystal structure of the DNA-QacR complex shows that the QacR dimers do not appear to interact directly, but both contribute to the required remodeling of the target DNA (Figure 17). This joint contribution to remodeling is thought to be the cause of the cooperative nature of QacR's interaction with DNA. This multimeric binding of proteins to DNA might not be the only example of such binding: another TetR family member, EthR, cooperatively binds as 4 dimers to an operator twice the size of QacR (*44*). Modeling of this EthR-DNA interaction indicates that the dimers will not interact directly. Thus the mechanism of cooperativity may be similar to the interaction of QacR with DNA (*41*). However, the mechanism of EthR-DNA binding has not been validated by an EthR-DNA complex structure.

The ligand-bound structures of most TetR members, including TetR, CprB, and EthR, indicate ligands bind in both subunits of the dimer. These protein ligand structures are inherently symmetrical. In fact, there is only a monomer in the asymmetric unit in one of the EthR-ligand structures; the other member of the dimer is related through the crystal symmetry (*41*). Though the CprB structure does not actually contain ligand it is thought that ligands would bind both subunits, as each subunit has an empty binding

pocket (*32*). QacR is the exception. It only binds a single ligand per dimer. This asymmetry in binding has been observed in all of the QacR structures solved, as well as in the stoichiometry of binding measured by ITC and equilibrium dialysis (*1*).

As there is little sequence homology in the ligand binding domain and ligand preferences for the TetR members are very different, it is not surprising that the binding pockets by the TetR family members are all different. TetR has a binding pocket orthogonal to the long axis of the protein, whereas the binding pockets of CprB and EthR run long and thin and parallel to this axis. The CprB binding pocket has a volume of 400 $Å^3$ approximately. Comparatively, the ligand-binding site for QacR is approximately 1100 $Å^3$, consisting of residues from almost every helix in the drug-binding subunit and as well as a few residues from the empty subunit.

The mechanism of induction from ligand binding is only known for TetR and QacR. EthR and CprB do not yet have the requisite sets of both DNA-bound and drug-bound complex structures to compare with one another. Tetracycline-Mg^{2+} binds to the TetR binding pocket between helices α9', α5 and, α7 and helices α4 and α6 (*38*). On one hand, the residues from the first set of helices do not move when comparing the positions of the residues before and after ligand binding. On the other hand, residues from helices α4 and α6 are displaced by 2 to 4 Å (Figure 18A). These residue movements are partly mediated by gross helical movements, which translate directly into movements of the DNA-binding domain. The distances between the recognition helices α3 and α3' move apart from 36.6 Å to 39.6 Å, which moves the recognition helices from the major groove and displaces the protein from the DNA (*38*).

In QacR, ligand binding involves a more complicated mechanism—not only gross helical movements but partial remodeling of the secondary structure of the protein. The residues 92 and 93 occupy the ligand-binding site in the DNA bound form of the protein. Ligand binding expels Y92 and Y93 from the binding pocket and triggers the reordering of residues T89-Y93 from random coil to α-helix (Figure 18B)(*1*). This transition lengthens the C-terminal end of α5 by one turn, and this extra turn pushes α6 away from the center of the dimer. Due to the tight interaction between α6 and α1, this extra turn on α5 also pushes the α1 of the DNA-binding domain away from the central axis of the protein (Figure 17C). This movement invokes a 9.1Å translation and 36.7° rotation of the DNA-binding domains compared to the DNA-bound conformation (*1*). As QacR binds drugs asymmetrically, the other subunit remains free of ligand and does not move directly due to drug binding. However, due to interactions between α6 of the ligand bound subunit and α6' of the ligand free subunit, conformational changes induced by ligand binding also occur in the ligand free subunit. The DNA binding portion of the ligand-free subunit moves out, but only with a 3.9 Å translation and a 18.3° rotation compared to the DNA-bound conformation (*1*). The resulting increase in the center-to-center distances of the recognition helices of the DNA binding domain is 11 Å, much greater than the 3 Å increase in distance measured in the TetR repressor (*1*). At this distance, the recognition helices are too far from each other to recognize the DNA, and thus QacR and its cognate DNA separate.

QacR: Ligand Binding

If there is one environmental toxin that pressured *S. aureus* to create QacR/QacA system for protection, it is difficult to determine its identity, as these proteins bind many

different molecules with similar affinity. This promiscuity is what makes this system an effective multidrug protection system and is one of the reasons these proteins are intriguing. QacR and QacA interact with over 30 different molecules of 11 different chemical classes, yet all the known ligands are all lipophilic cations (Figure 19). It is interesting to explore what allows so many different molecules interact with one binding pocket and how the binding pocket selects for positive charge.

The QacR binding pocket consists of two distinct but connected smaller pockets. They are each named for the representative ligands that bind in each pocket: R6G and Et. These pockets overlap, as R6G and Et binding to QacR is mutually exclusive. However, proflavin, which binds in the R6G minipocket, and ethidium bind to QacR simultaneously (*8*). The different ligand binding characteristics of each minipocket and the connectivity between them allows for the wide variety of structures and sizes of ligands that bind to QacR.

QacR uses various types of residues to create a multifaceted binding pocket that can change shape in order to accommodate these disparate molecules. Foremost, the binding pocket is lined primarily with aromatic residues. These aromatic residues can form many different types of interactions with the ligands in the binding pocket. These interactions include π-π interactions with aromatic ligands, cation-π interactions with positively charged regions of ligands, and van der Waals interactions with any ligand. Furthermore, the tyrosyl hydroxyls can also provide hydrogen bonds to the ligands. The aromatic residues in the pocket, namely the tyrosines, can also reorient themselves to accommodate ligands of various shapes. Tyrosines 103 and 123 shift position to

accommodate and maximize binding potential of various drugs observed in complexes with QacR (Figure 20A)(*1*).

Similarly, hydrophobic residues also provide van der Waals interactions to any ligand in the pocket. Methionine 116, isoleucine 100, and other hydrophobic residues interact with ligands bound to any side of the binding pocket (*1*). As the known substrates thus far are hydrophobic ligands, it is not surprising that the binding pocket is hydrophobic. However, hydrophilic residues are scattered through the binding pocket to provide hydrogen-bond interactions for the correctly oriented ligand.

A particularly notable aspect of the binding pocket is the handful of glutamates evenly distributed in the pocket (Figure 1 and 20B). Of the five glutamates in the binding pocket, E90 resides in the R6G-binding pocket and comes within 4 Å of R6G, within 3.3 Å of MG, and within 3.7 Å of Et (*1*). Glutamate 120 interacts with Et at a distance of 4 Å. Malachite green also comes within 3.3 Å of E120, and one end of the bivalent ligand Dq comes within 4.8 Å (*1*). Dequalinium also interacts with E57 and E58 jointly on the other end of this long ligand with distances of 4.1 Å and 6.2 Å, respectively (*1*). Recently, it was discovered that although the bivalent molecule hexamidine binds similarly to Dq, its shorter homologue pentamidine binds in a unique way. It interacts with E63, a residue that resides on the surface of QacR just on the outside of the binding pocket (Figure 2)(*7*). On the other end of pentamidine's alkyl chain the benzamidine head-group of the pentamidine interacts with the hydroxyl of Y127 rather than another glutamate (*7*). With the exception of the one end of the bivalent ligand pentamidine that is complemented by Y127 all of the cations in the pocket directly interact with glutamates.

The greatest contributions to ligand binding appear to be from the aromatic residues and charge complementarity of the glutamates in the pocket. The aromatic residues can provide a number of types of interactions with the ligands, including hydrogen bonding with the hydroxyl groups of the tyrosines. The aromatic residues as well as a few hydrophobic residues provide the majority of the interactions in the binding pocket. It appears, however, that the glutamate residues play an important role in binding with charge complementarity to the cationic ligands.

Complementary Charge Interactions in Ligand Binding

The literature indicates a number of examples concerning the importance of complementary charge interactions in ligand-protein interactions. The protein that find these complementary charge interactions to be important include both include both membrane bound and soluble proteins in prokaryotes and eukaryotes (*45-47*).

The structures of BmrR, indicate charge complementation between the ligand and the protein is important. Substitution of the buried E253 to alanine leads to a modest four-fold decrease in binding (*30*). Though this decrease is not dramatic, the small change in affinity may due to a repositioning of the drug in the pocket to a position closer to D47. Substitution of D47 to a neutral residue has no effect on binding, but when combined with the E254A/D substitutions, there is a greater than 10 fold decrease in R6G affinity (Brennan Laboratory, unpublished).

Similarly, the wide-spectrum MFS transporter, MdfA, has only one charged amino acid, glutamate 26, located in the intramembrane portion of the protein (*48*). The charge at this position is absolutely necessary for recognition and transport of positive ligands. Cells expressing the multidrug transporter with a lysine or alanine substitution at

residue 26 can no longer protect the cells from positively charged ligands such as Et or benalkonium (*48*). However, cells expressing MdfA with a substitution of aspartate at this position show resistance to the positively charged ligands (*48*). Requirement of an acidic residue at this position is specific for ligand recognition rather than coupling ligand transport to the proton gradient, as substitution of E26 with lysine allows proton coupling for chloramphenicol transport (*48*).

Besides MdfA, many other drug transporters that transport lipophilic cations across the membrane contain an acidic group at a similar position (Figure 21)(*49*). In QacA, another MFS transporter, this corresponds to a glutamate 34. All the known members of the Smr family, which include the lipophilic cation transporters Smr (QacC), QacG, QacH, and QacJ, also contain a glutamate in a corresponding position (*50*). Similar to MdfA, this glutamate is the only charged residue in the transmembrane domains of the SMR family members.

The conserved location of this acidic residue in these transporters also suggests that the positional context of the negative charge in the protein is important as well. In fact, search for a second site suppressor for an E26T substitution in MdfA yielded only carboxylic acid substitutions at position 335. As cross-linking studies between positions 26 and 335 places them on opposite sides of a large pocket, it is understandable that glutamate at position 335 may substitute for 26 if the charge is important for binding ligands. This theory is supported by the fact that transport is not restored for all of the positive ligands, as it would be likely, given the constraints of a binding pocket, that not all ligand would be able to reorient itself to align with residue 335.

The ligand specificity of QacA also seems to be determined by a charged residue. Not only does QacA have a conserved negative charge in the transmembrane region at residue 34, but it also has an aspartate at position 323. This second aspartate is absent in QacB, which can only transport monovalent cations, suggesting that an aspartate at this position could be important for interacting with the second positive charge of a bivalent cation. Random mutagenesis of QacB yielded protection from bivalent cations only when the mutant proteins contained a negatively charged residue at position 322 or 323 (*11*).

As QacA and QacR interact with an almost identical set of ligands, it is possible that they use similar mechanisms for ligand recognition. Research suggests that buried negative charges in the transmembrane domains 1 and 10 may confer the ability to bind negatively charged ligands in QacA; thus, the four negative charges in the binding pocket of QacR may act to neutralize the charges of the cationic drugs. The drugs in the pocket all bind proximal to at least one, if not two or three, glutamates. It is likely that these charges play a role in ligand recognition. I hypothesize that a substitution of a glutamates with non-charged residues such as alanine and glutamine will lower the affinity of ligands that interact with the glutamate in the wild-type protein.

The structural characterization and biochemical analysis of ligand binding of the substituted QacR proteins E90Q-QacR and E120Q-QacR directly assess the roles of these glutamates in ligand binding. Furthermore, using structural characterization and biochemical analysis of ligand binding studies, the diamidines DB75 and DB359 probe the WT-QacR binding pocket to determine if either of these molecules binds similarly to previous ligands or unique ways that may include other glutamates.

Figure 1. Multidrug binding to QacR. A. R6G binding proximal to E90. B. Ethidium binding proximal to E120. C. Dequalinium binding proximal to E57, E58, and E120. D. Crystal violet binding proximal to E90 and E120. E. Malachite Green binding proximal to E90 and E120. F. Berberine binding proximal to E57 and E58.

Reproduced from Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A., and Brennan, R. G. (2001) Structural mechanisms of QacR induction and multidrug recognition, *Science 294*, 2158-2163.

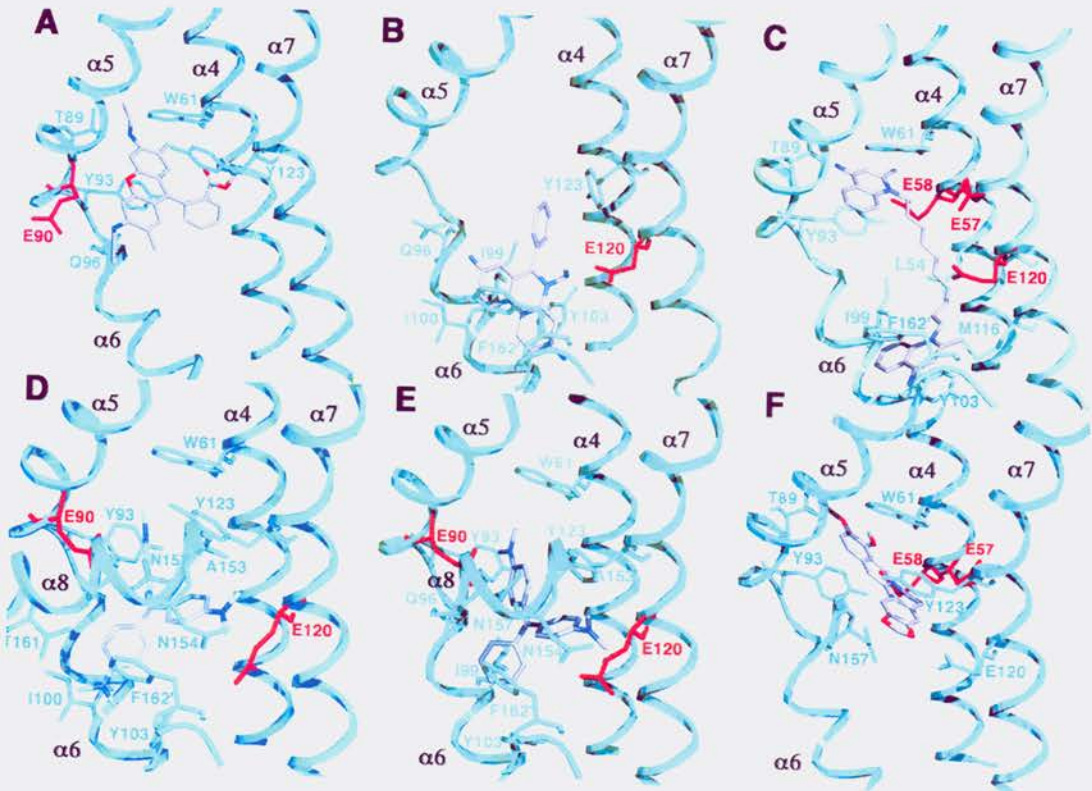
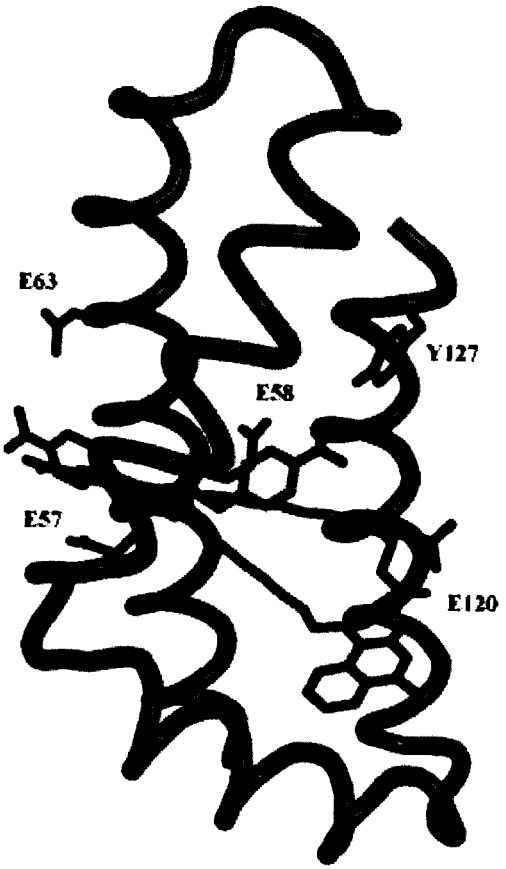

A
α5
α4
α7
W61
T89
Y93
Y123
E90
Q96
α6
B
α5
α4
α7
Y123
Q96
I99
E120
I100
Y103
F162
α6
C
α5
α4
α7
W61
T89
E58
E57
Y93
L54
E120
I99
F162
M116
α6
Y103
D
α5
α4
α7
W61
E90
Y93
Y123
α8
N154
A153
E120
T161
N154
I100
F162
Y103
α6
E
α5
α4
α7
E90
α8
E120
α6
F
α5
α4
α7
W61
T89
E58
E57
Y93
Y123
N157
E120
α6

Figure 2. Pentamidine and hexamidine binding to QacR. Pentamidine (light green) interacts with E63 and the hydroxyl from Y127. Hexamidine (purple) and dequalinium (pink) interact with E57, E58, and E120.


E63
Y127
E58
E57
E120

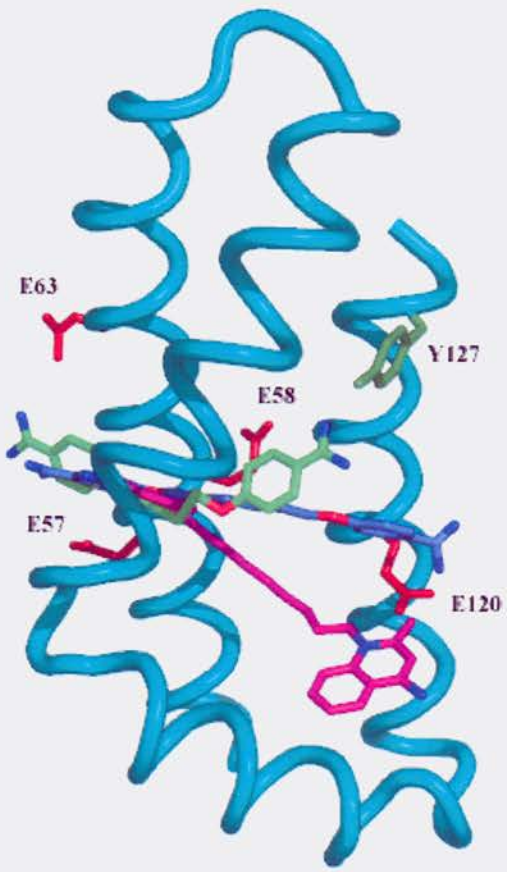
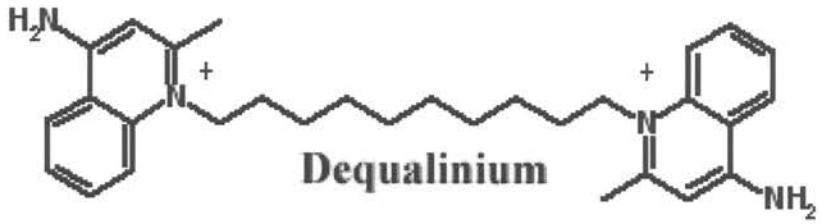
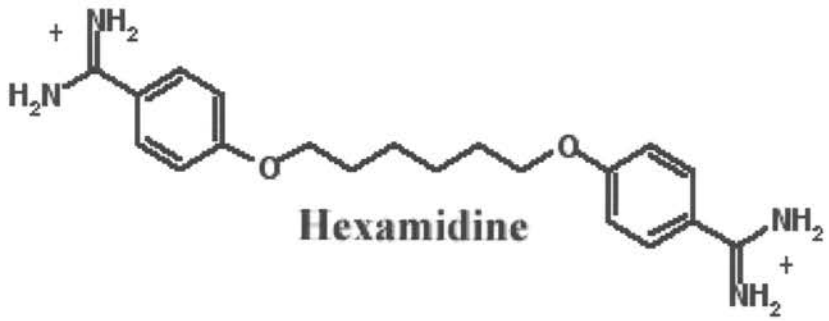

E63
Y127
E58
E57
E120

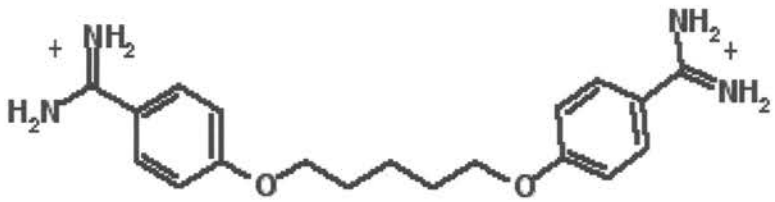
Figure 3. Comparison of diamidines.



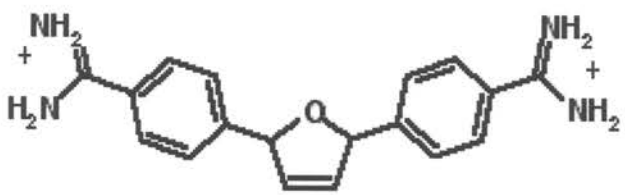
Dequalinium



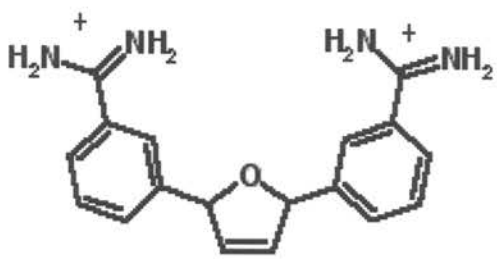
Hexamidine



Pentamidine



DB75



DB359

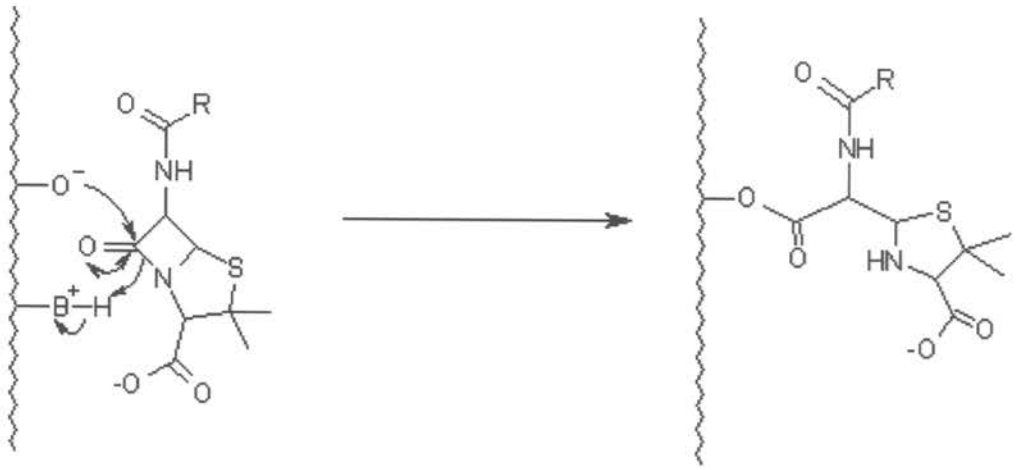
Figure 4. Mechanisms of action of β-lactam inhibition and β-lactam resistance. A. A base catalyzed nucleophilic attack by the penicillin binding protein on the β-lactam fixes the inhibitor to the active site. B. An example of the mechanism of a type 2 β-lactamase adding water to a β-lactam ring to open the ring and render the drug ineffective.

Figure adapted from the following websites:

www.chemsoc.org/exemplarchem/entries/2002/stanley/06_Mechanism/Mechanism.htm

www-mitchell.ch.cam.ac.uk/gemma/ACS0309-MACiE.ppt

A


O
R
NH
O⁻
O
N
S
B⁺
H
-O
O
O
R
NH
O
O
S
HN
O
-O

B

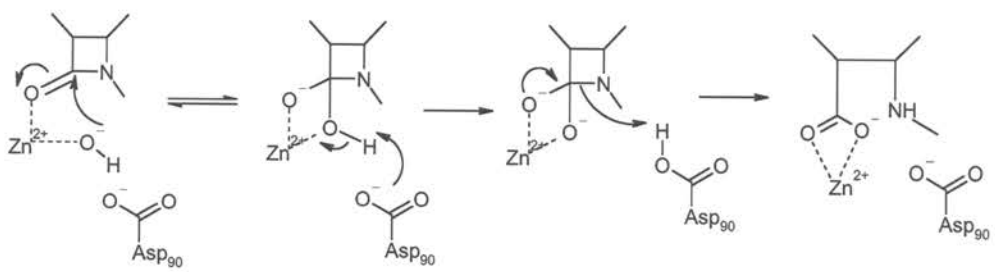

N
O
Zn2+
O
H
O⁻
O
Asp90
N
O⁻
O
Zn2+
O
H
O⁻
O
Asp90
N
O⁻
O
Zn2+
O⁻
H
O
O
Asp90
NH
O
O⁻
Zn2+
O⁻
O
Asp90

Figure 5. Topology of the 14 transmembrane domain major facilitator superfamily proteins QacA and QacB. The D323G mutation is the only difference between the amino acid sequences of these two transporters.

Modified from Paulsen, I. T., Brown, M. H., Littlejohn, T. G., Mitchell, B. A., and Skurray, R. A. (1996) 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.

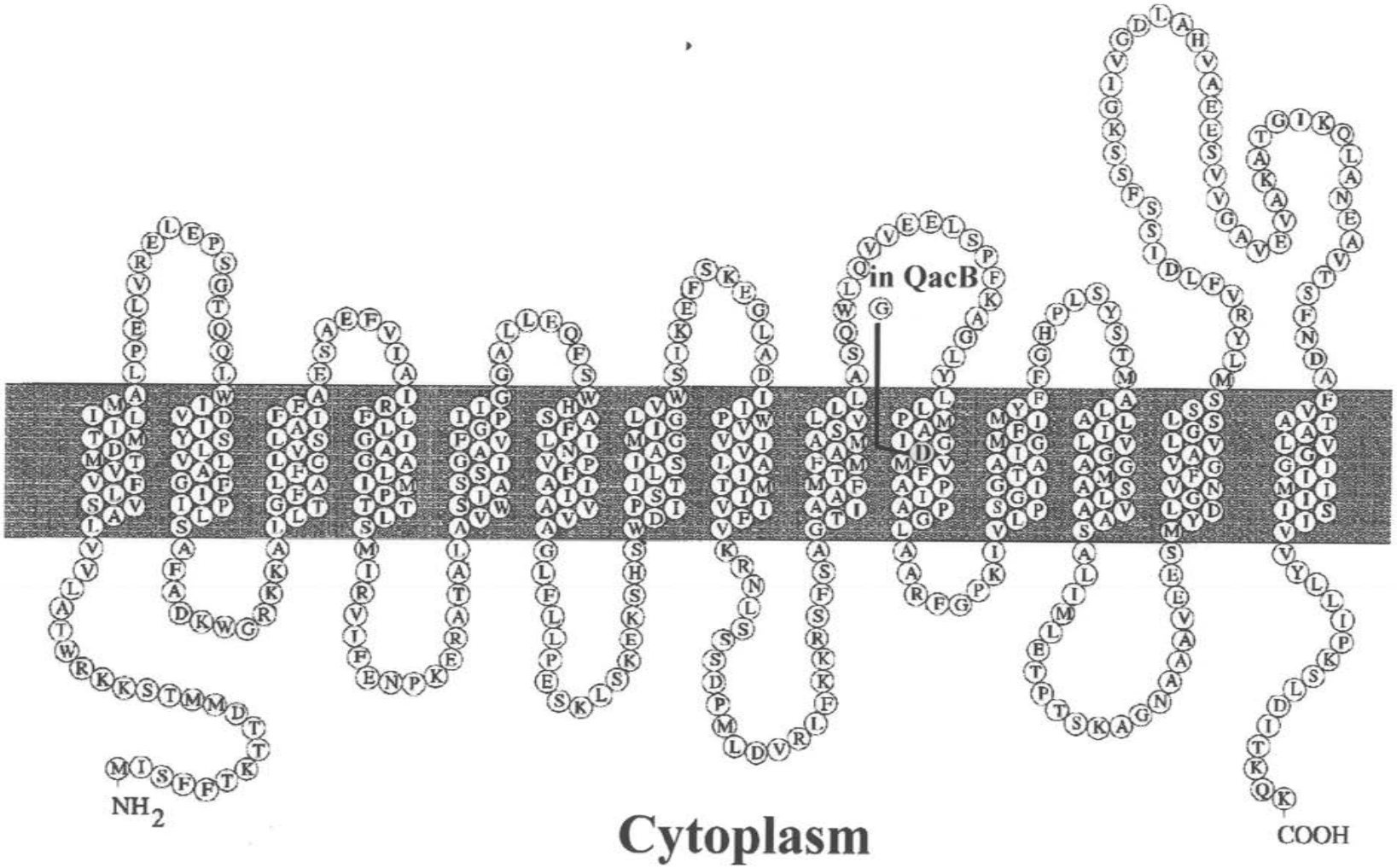

in QacB
NH2
COOH
Cytoplasm

Figure 6. Regulatory controls of the *E. coli acrAB* and *tolC* gene expression. The AcrAB-TolC transport complex extrudes drugs across the bacteria's outer (OM) and cytoplasmic (CM) membranes. The expression of AcrA and AcrB is prevented locally by the self-regulating repressor AcrR (green). However, expression of the components of the AcrAB-TolC transport complex occurs by the global regulatory proteins MarA, SoxA, and Rob (all in purple). Any one of which can activate transcription by binding the marbox (cyan). The expression of MarA is regulated by MarR (red), which is a repressor of MarA, MarB, and MarR expression that binds to the operator site *marO*. Induction of MarR occurs on drug (diamond) binding and MarA is expressed. SoxS is regulated by SoxR (pink), which is induced by radical oxygen.

Figure adapted from Grkovic, S., Brown, M. H., and Skurray, R. A. (2002) Regulation of bacterial drug export systems, *Microbiol Mol Biol Rev 66*, 671-701.

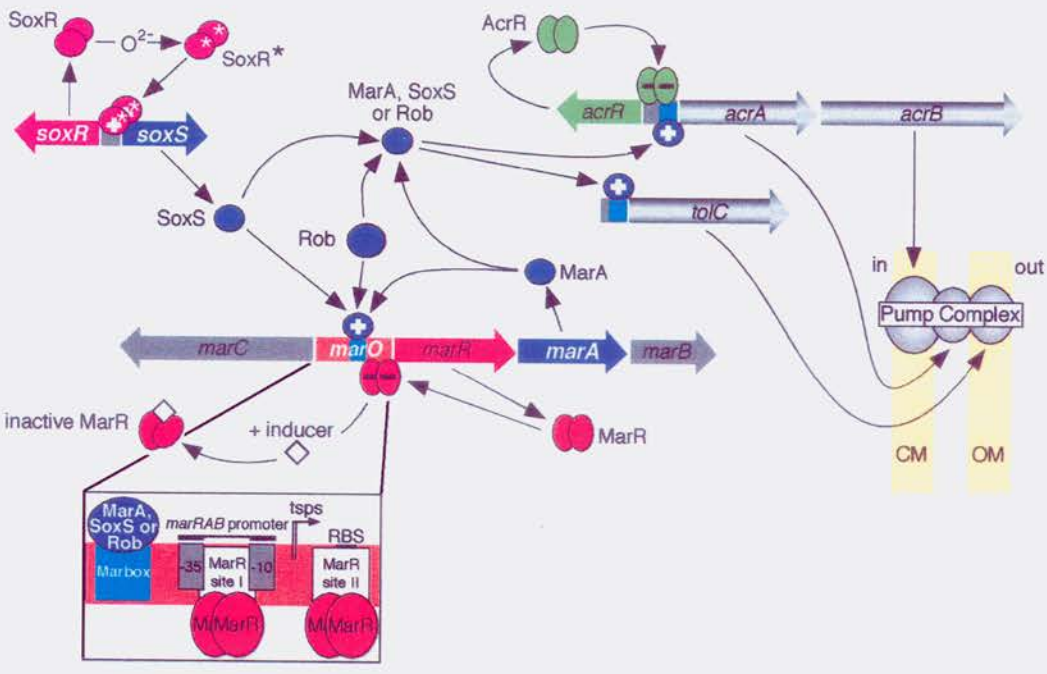

SoxR
O^{2-}
SoxR*
soxR
soxS
SoxS
MarA, SoxS or Rob
Rob
MarA
AcrR
acrR
acrA
acrB
tolC
in
out
Pump Complex
CM
OM
marC
marO
marR
marA
marB
inactive MarR
+ inducer
MarR
MarA, SoxS or Rob
marRAB promoter
tsps
RBS
Marbox
-35
MarR site I
-10
MarR site II
M(MarR)
M(MarR)

Figure 7. Crystal structures of Rob and MarA bound to DNA. In this family, the regulator binds as a monomer to DNA with two HTH DNA binding domains. The most significant difference between these structures is that Rob has a ligand-binding domain where as MarA does not. Furthermore, in these structures MarA inserts its DNA binding domains into successive major grooves where as Rob only inserts one DNA binding domain and the other interacts with the backbone of the DNA.

Reproduced from Kwon, H. J., Bennik, M. H., Demple, B., and Ellenberger, T. (2000) Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA, *Nat Struct Biol 7*, 424-430.

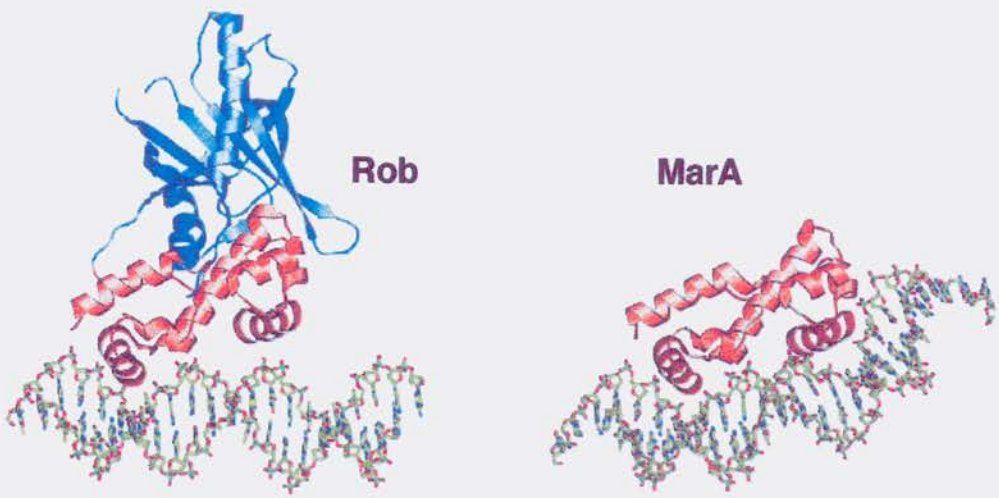
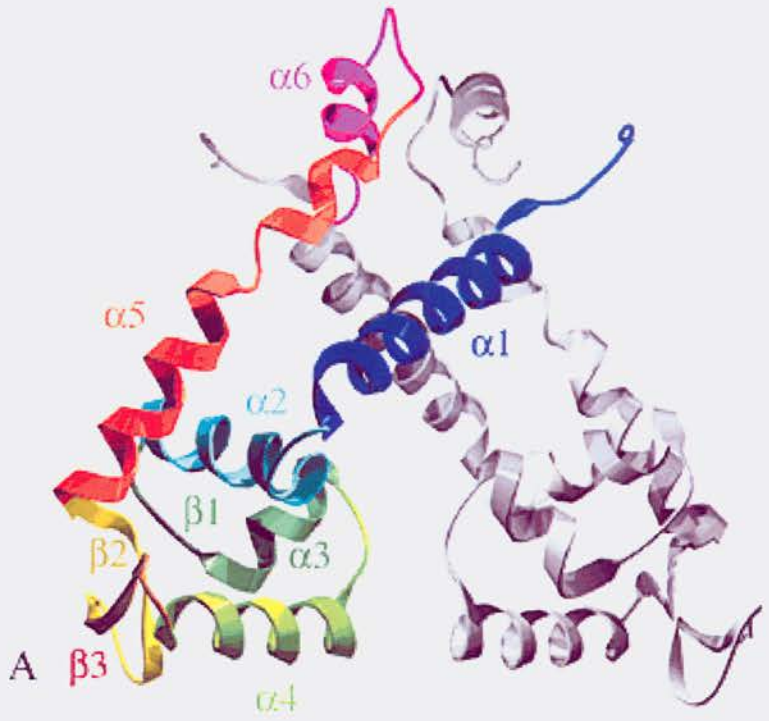

Rob
MarA

Figure 8. Apo and ligand bound structures of MarR. A. MarR is a homodimer, each subunit of which consisting of 6 α-helices and a small 3 stranded β-sheet with the over all structure of an equilateral triangle. B. MarR binds ligands symmetrically, one ligand in each monomer.

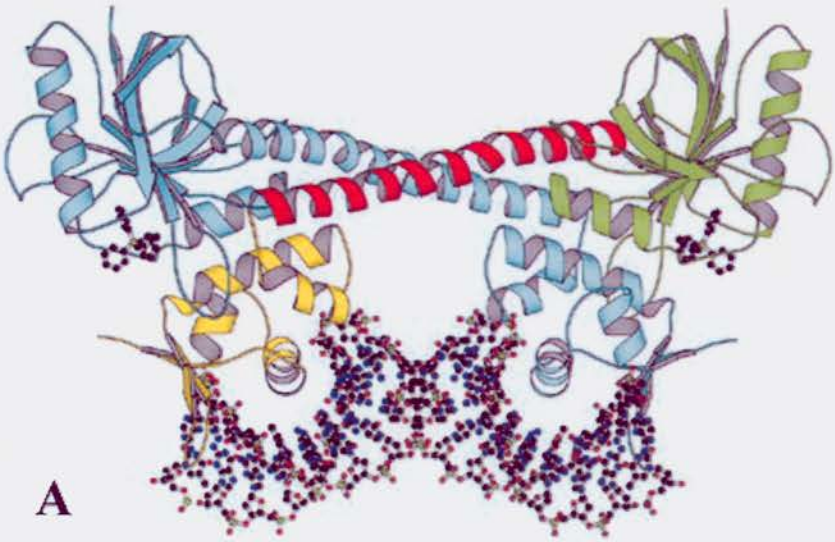
Reproduced from Alekshun, M. N., Levy, S. B., Mealy, T. R., Seaton, B. A., and Head, J. F. (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution, *Nat Struct Biol 8*, 710-714.


α6
α5
α1
α2
β1
β2
α3
A
β3
α4

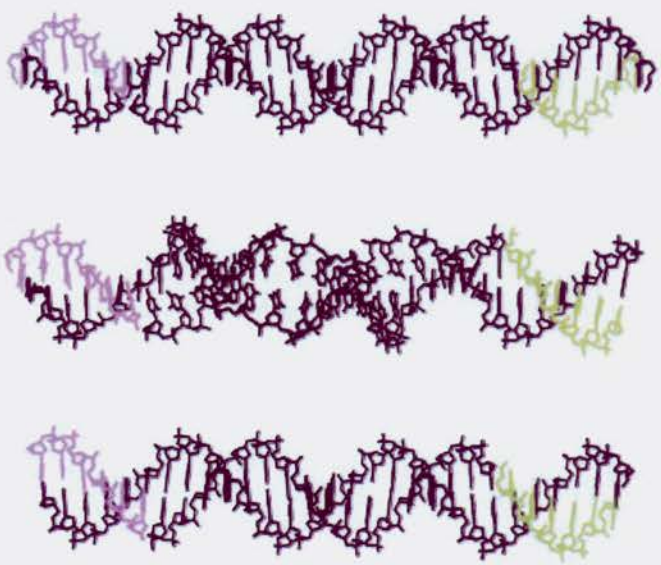

B

Figure 9. DNA activation by drug binding to BmrR. A. BmrR is bound to DNA and tetraphenylphosphine. The distortion of the DNA that is created on drug binding is apparent. B. The top duplex of DNA illustrates the 19 bp operator of BmrR with the -10 and -35 hexamers of the promoter in green and purple. The -10 and -35 hexamers are on opposite sides of the promoter. The bottom strand illustrates the canonical 17 bp operator where the -10 and -35 hexamers are on the same side of the DNA, aligned for RNA polymerase binding. The middle strand illustrates how BmrR remodels the DNA on drug binding and orients the -10 and -35 hexamers for RNA polymerase binding.

Reproduced from Heldwein, E. E., and Brennan, R. G. (2001) Crystal structure of the transcription activator BmrR bound to DNA and a drug, *Nature 409*, 378-382.



A



B

Figure 10. BmrR regulates the expression of Bmr. Bmr is a multidrug transporter that harnesses the proton gradient across the bacterial cytoplasmic membrane (CM) in order to efflux drugs from the bacteria. The global activator Mta (purple) or BmrR (red) activate the expression of Bmr. BmrR is constitutively bound to its promoter (small white arrow) and on drug binding remodels the DNA to promote RNA polymerase binding and expression Bmr.

Reproduced from Grkovic, S., Brown, M. H., and Skurray, R. A. (2001) Transcriptional regulation of multidrug efflux pumps in bacteria, *Semin Cell Dev Biol 12*, 225-237.

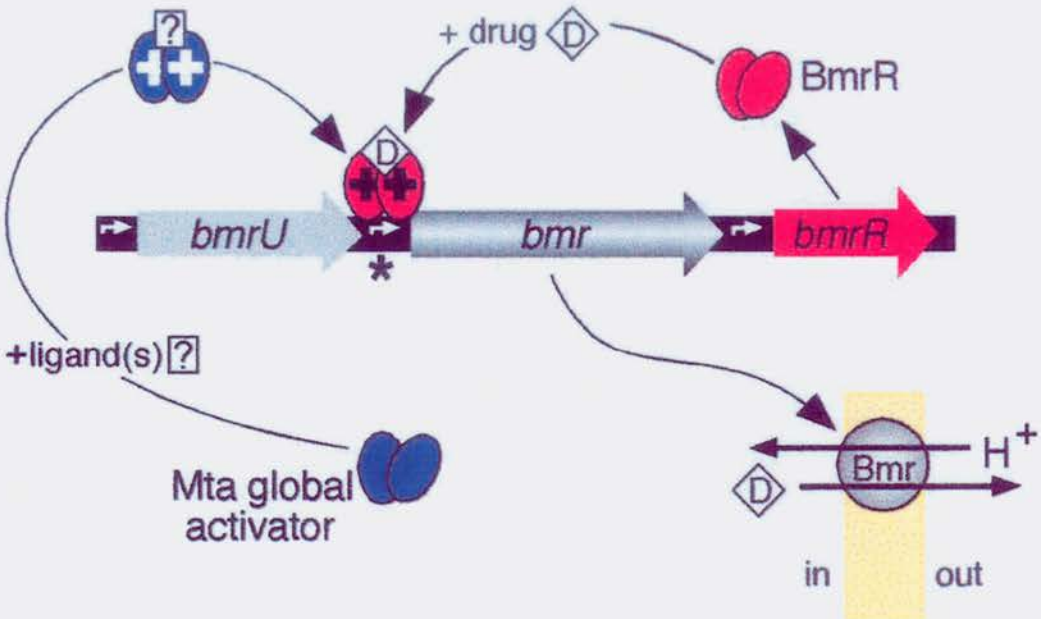
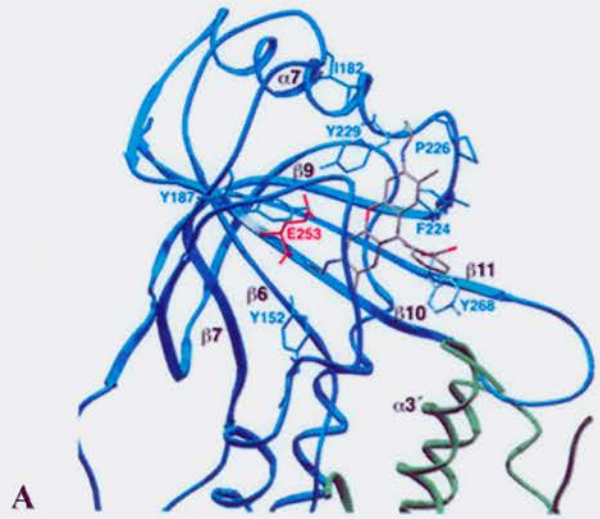
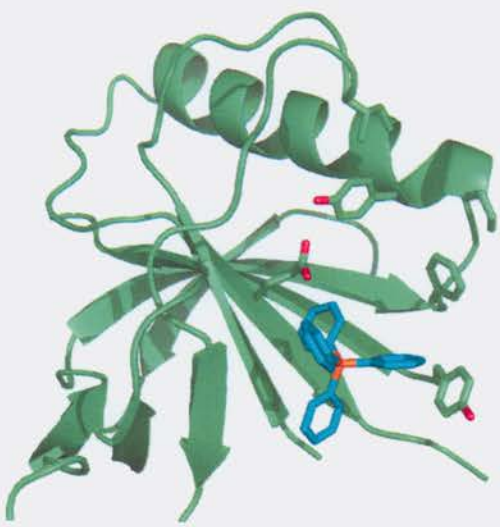

?
+ drug
D
BmrR
D
bmrU
bmr
bmrR
*
+ligand(s)?
Mta global
activator
Bmr
H+
D
in
out

Figure 11. Comparison of the BmrR-R6G and BRC-TPP complex structures. A. BmrR-R6G. B. BRC-TPP

Panel A was reproduced from Newberry, K. et al. EMBO in press



A



B

Figure 12. Movement of Y152 on ligand binding to BmrR

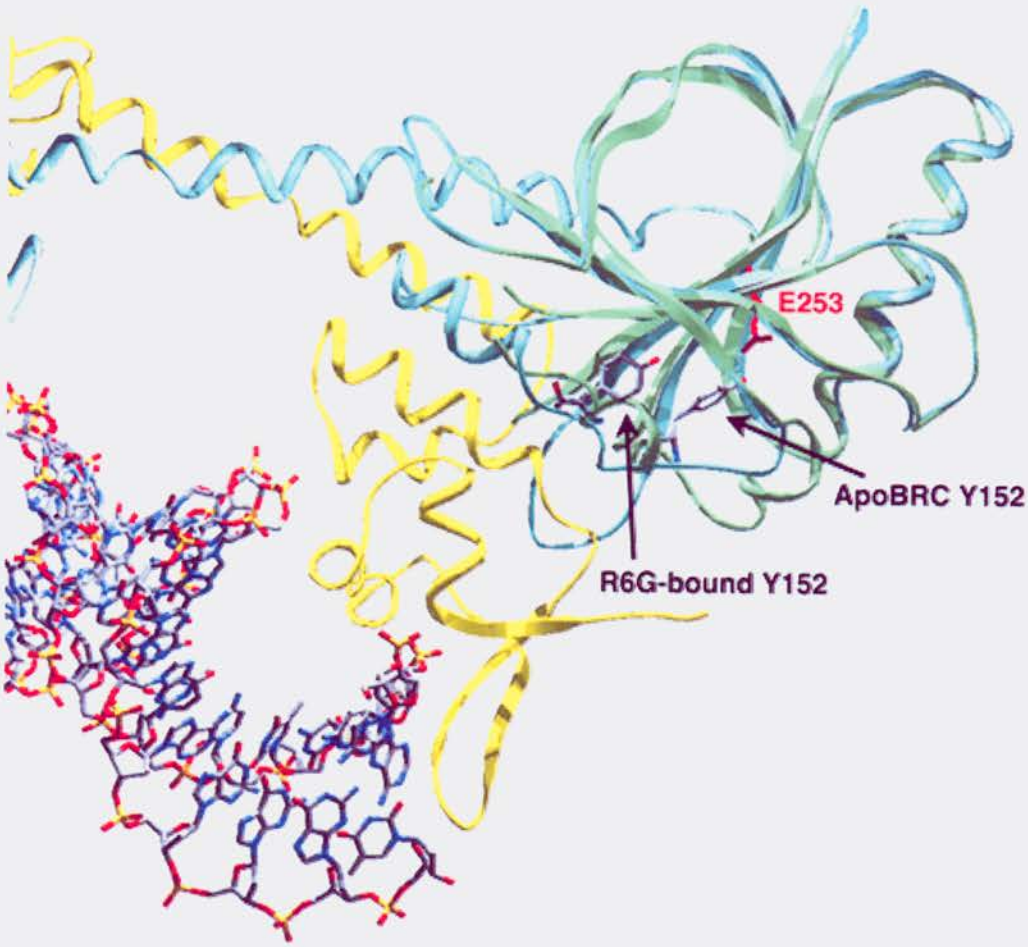

E253
ApoBRC Y152
R6G-bound Y152

Figure 13. TetR repression system. The two TetR operators, O_1 and O_2, overlap with the promoters for TetA expression, P_{tetA}, and the promoters for TetR expression, P_{tetR1} and P_{tetR2}. The genes are tightly expressed until TetR binds tetracycline-Mg^{+2} and is induced and more TetR and TetA are expressed. TetA traffics to the cytoplasmic membrane and effluxes tetracycline-Mg^{+2} out of the cell.

Reproduced from Grkovic, S., Brown, M. H., and Skurray, R. A. (2002) Regulation of bacterial drug export systems, *Microbiol Mol Biol Rev 66*, 671-701.

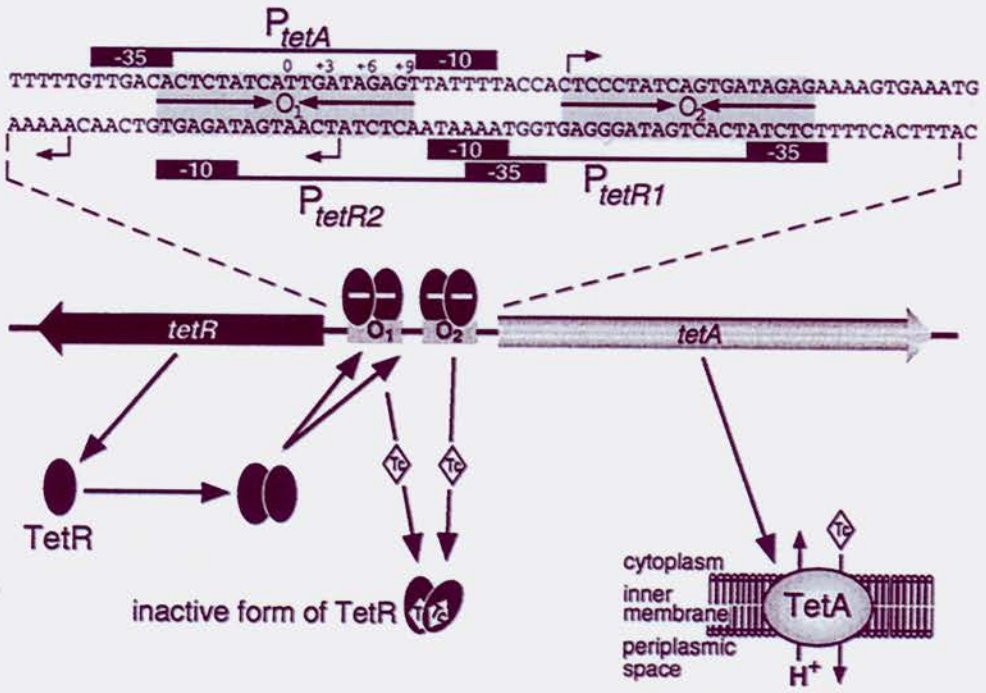

P tetA
-35
0 +3 +6 +9
-10
TTTTTGTTGACACTCTATCATTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAATG
O1
O2
AAAAACAACTGTGAGATAGTAACTATCTCAATAAAATGGTGAGGGATAGTCACTATCTCTTTTCACTTTAC
-10
-35
-10
-35
P tetR2
P tetR1
tetR
O1
O2
tetA
TetR
Tc
Tc
inactive form of TetR
Tc
cytoplasm
inner membrane
periplasmic space
TetA
H+

Figure 14. Regulation of QacA expression by QacR.

Reproduced from Grkovic, S., Brown, M. H., and Skurray, R. A. (2002) Regulation of bacterial drug export systems, *Microbiol Mol Biol Rev 66*, 671-701.

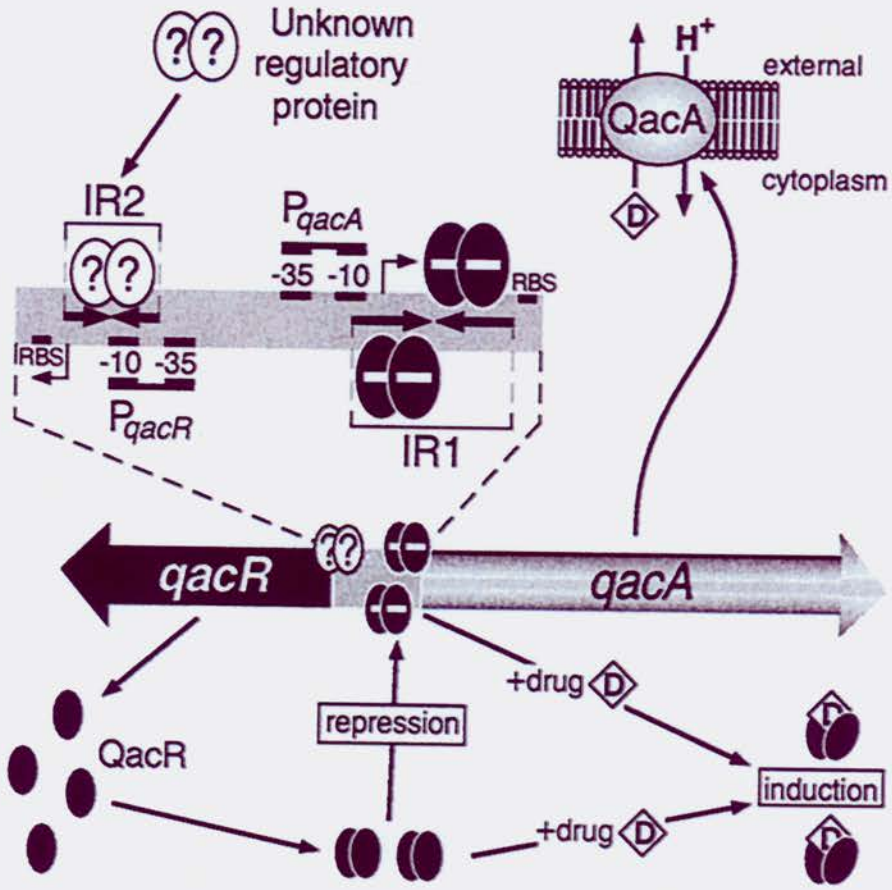

Unknown
regulatory
protein
IR2
P_{qacA}
-35 -10
RBS
IRBS
-10 -35
P_{qacR}
IR1
H^+
external
QacA
cytoplasm
D
qacR
qacA
+drug
repression
QacR
induction
+drug

Figure 15. QacR operator and QacR DNA binding. A. QacRs operator, IR1, consists of a double inverted repeat the half sites of which are separated by 6 pb. The red box indicates which bases can be switched to strengthen repression. B. QacR binds to IR1 as a dimer of dimers.

Reproduced from Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A., and Brennan, R. G. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR, *EMBO J 21*, 1210-1218.

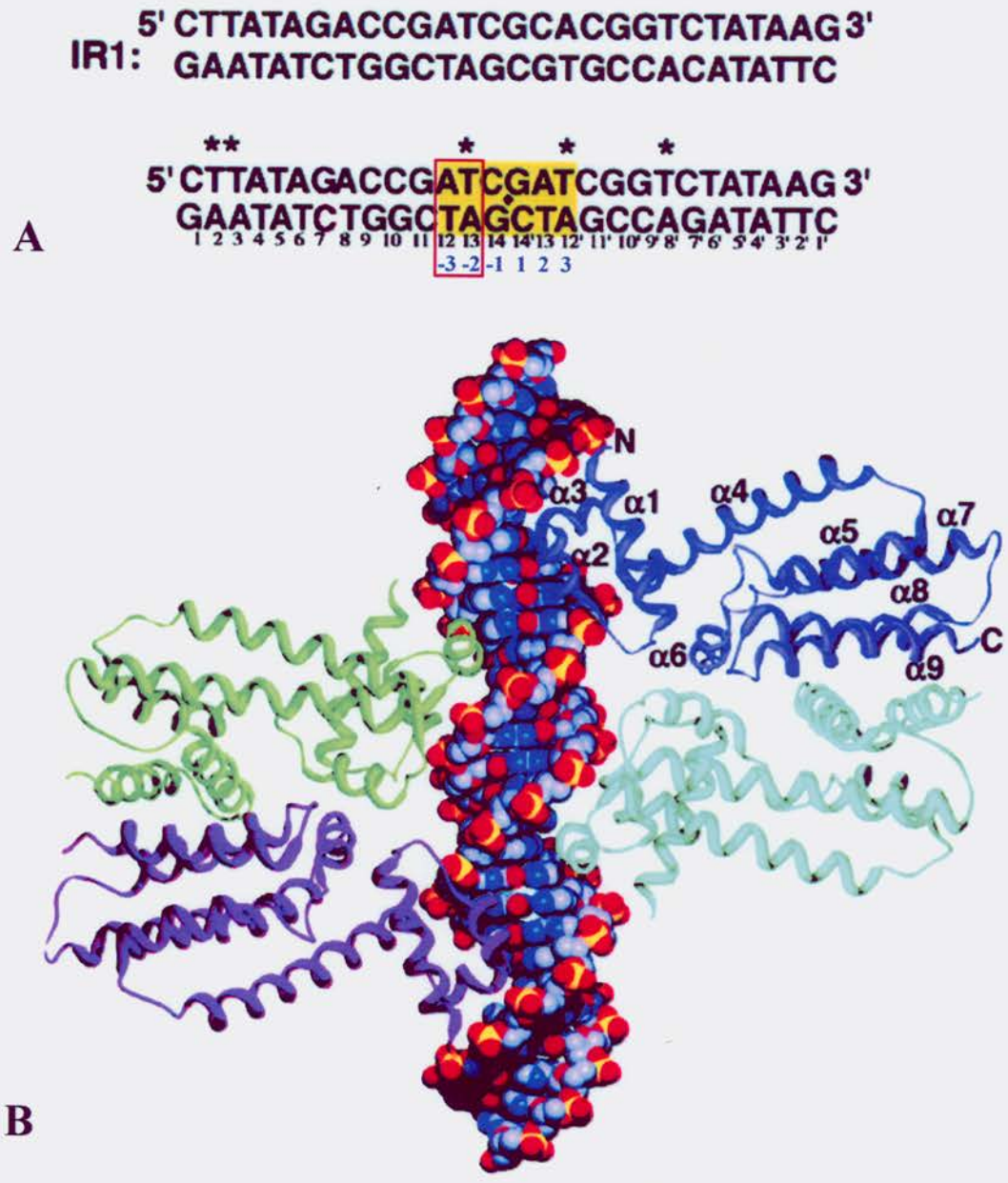
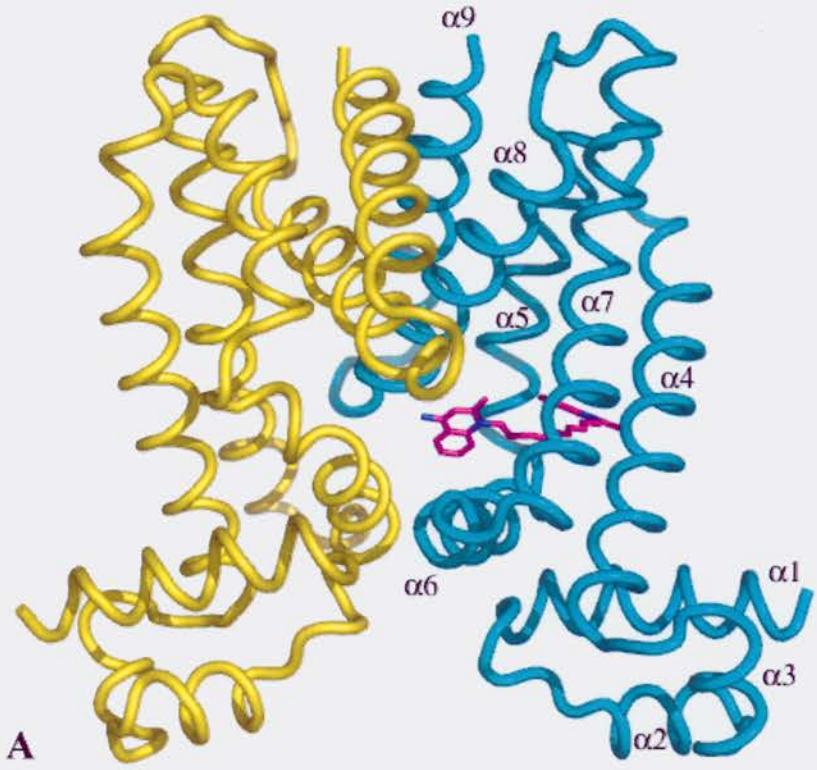

IR1:
5' CTTATAGACCGATCGCACGGTCTATAAG 3'
GAATATCTGGCTAGCGTGCCACATATTC
A
5' CTTATAGACCGATCGATCGGTCTATAAG 3'
GAATATCTGGCTAGCTAGCCAGATATTC
1 2 3 4 5 6 7 8 9 10 11 12 13 14 14' 13' 12' 11' 10' 9' 8' 7' 6' 5' 4' 3' 2' 1'
-3 -2 -1 1 2 3
B
N
α1
α2
α3
α4
α5
α6
α7
α8
α9
C

Figure 16. Representative structures of TetR repressor family members. A. QacR bound with Dq. QacR consists of 9 α-helices and only one ligand binds per dimer. B. TetR bound to tetracycline-Mg^{+2}. TetR consists of 10 α-helices and two ligands bind per dimer.


α9
α8
α7
α5
α4
α6
α1
α3
α2
A

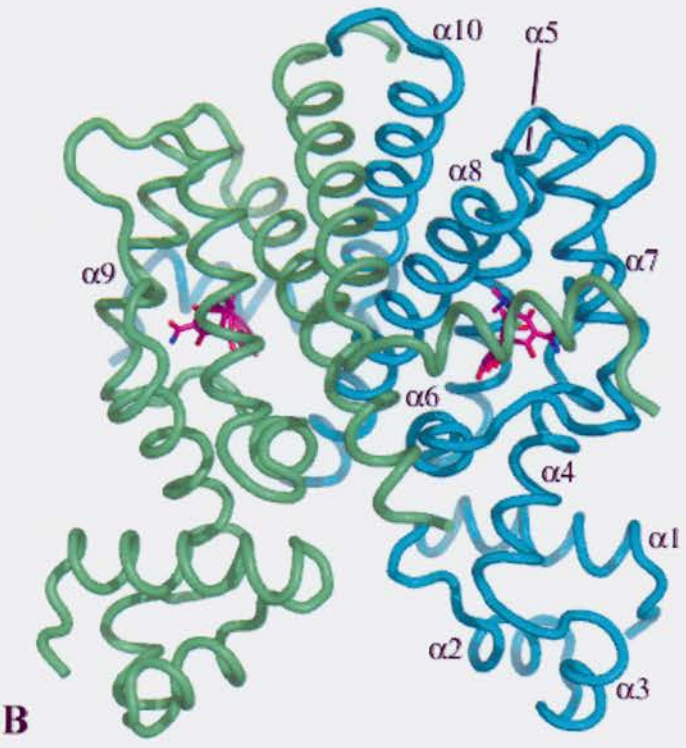

α10
α5
α8
α9
α7
α6
α4
α1
α2
α3
B

Figure 17. IR1 is undertwisted by QacR binding.

Reproduced from Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A., and Brennan, R. G. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR, *EMBO J 21*, 1210-1218.

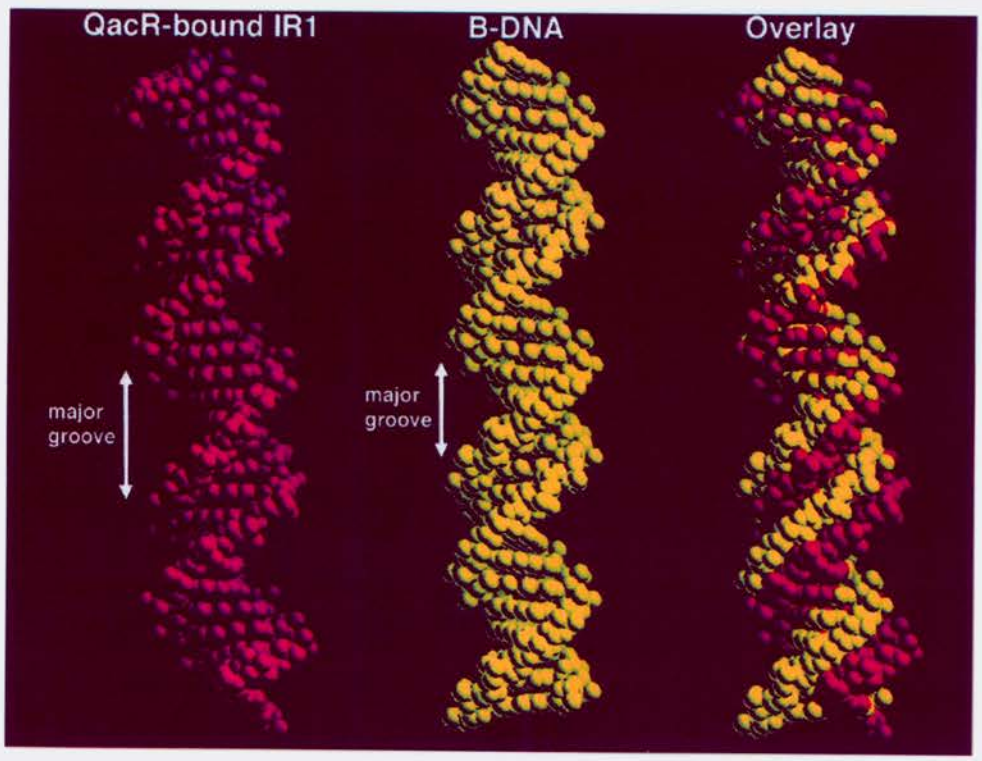
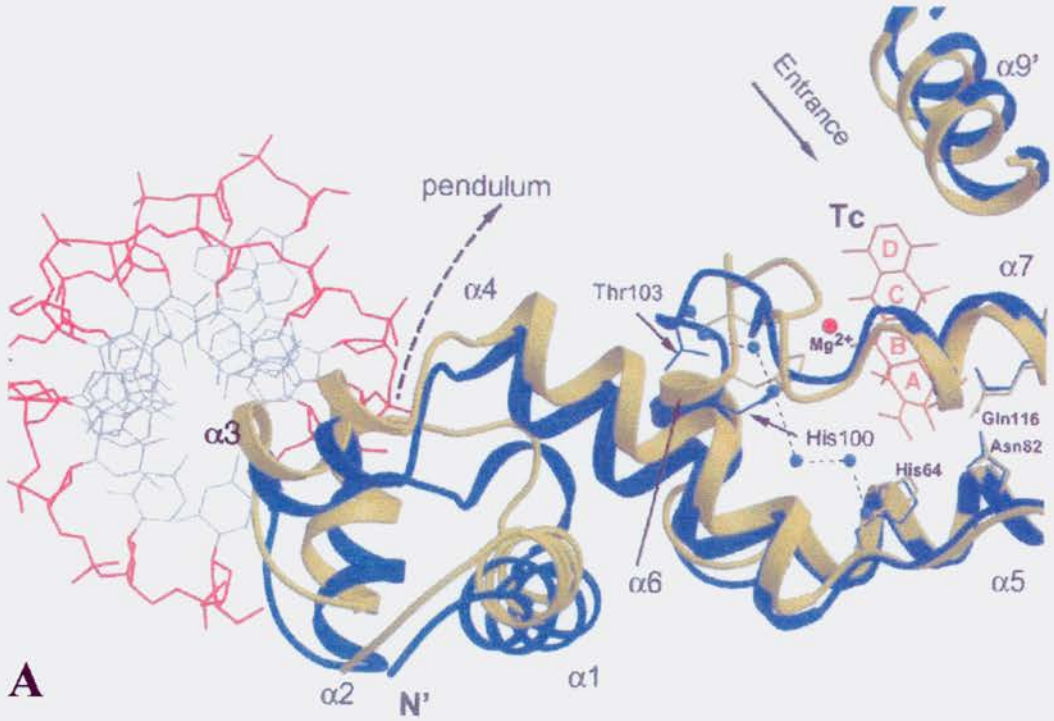

QacR-bound IR1
B-DNA
Overlay
major groove
major groove

Figure 18. Comparison of TetR and QacR induction mechanisms. A. Tetracycline-Mg^{+2} binding pulls the DNA binding domain outwards by concerted ridged body helical motions. B. Tyrosines 92 and 93 from the WT-QacR-DNA (yellow) complex structure occupy the ligand-binding pocket. These residues are expelled on R6G binding causing a coil-to-helix transition at the end of α5. C. The coil-to-helix transition forces α6 away from the center of the protein and through interhelical interactions forces the DNA binding domains away from the center of the protein.

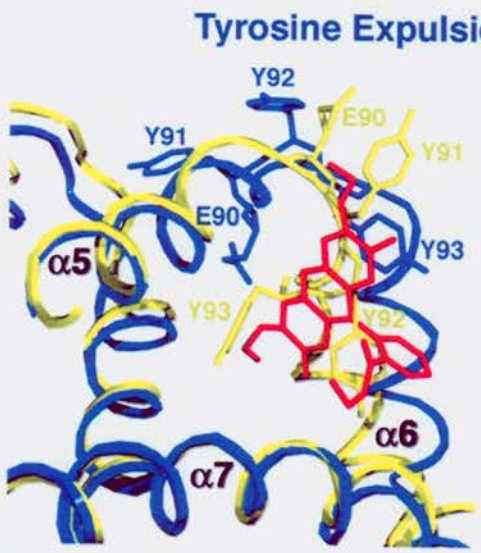
Reproduced from Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system, *Nat Struct Biol 7*, 215-219.

and

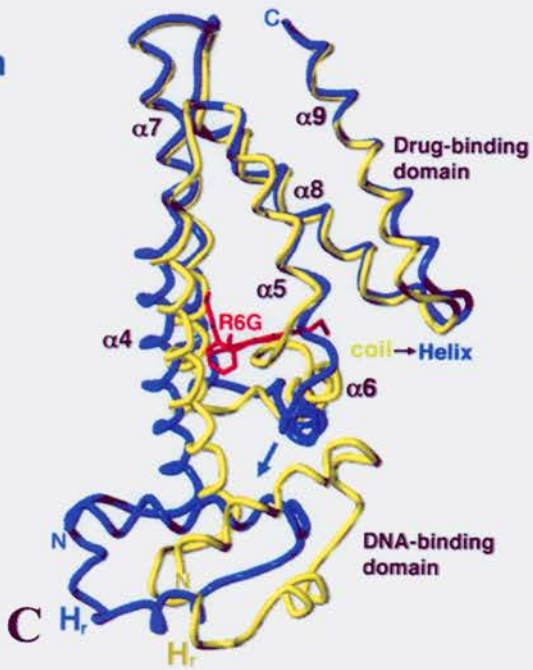
Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A., and Brennan, R. G. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR, *EMBO J 21*, 1210-1218.


α9'
Entrance
pendulum
Tc
α4
Thr103
α7
D
C
B
A
Mg2+
α3
His100
Gln116
Asn82
His64
α6
α5
α2
N'
α1

A


Tyrosine Expulsion
Y92
E90
Y91
Y91
E90
α5
Y93
Y93
Y92
α6
α7

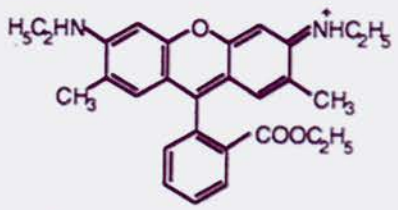
B


C
α7
α9
Drug-binding domain
α8
α5
α4
R6G
coil→Helix
α6
N
DNA-binding domain
N
Hr
Hr

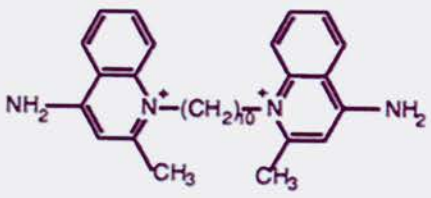
C

Figure 19. QacR ligands.

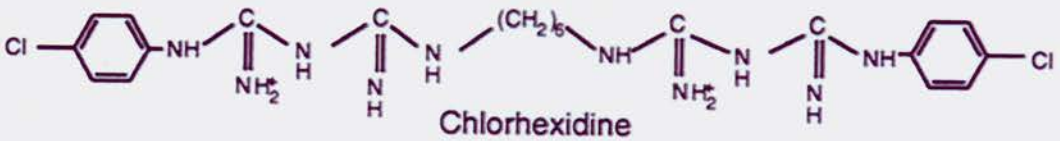
Reproduced from Grkovic, S., Brown, M. H., and Skurray, R. A. (2002) Regulation of bacterial drug export systems, *Microbiol Mol Biol Rev 66*, 671-701.



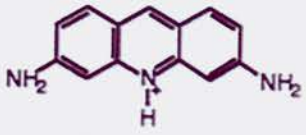
Rhodamine 6G



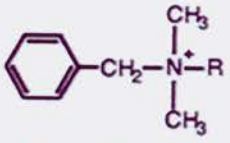
Dequalinium



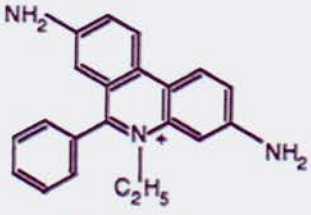
Chlorhexidine



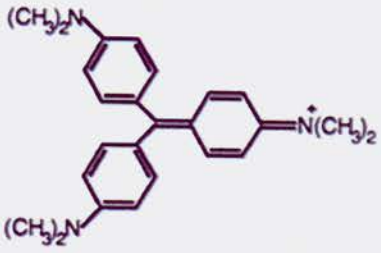
Proflavine



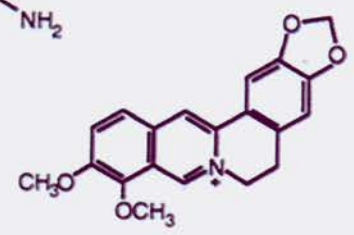
Benzalkonium



Ethidium

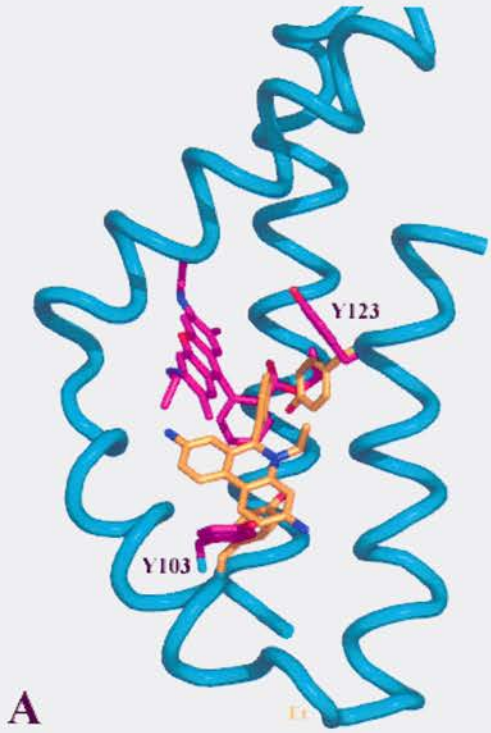


Crystal violet



Berberine

Figure 20. QacR residues that are hypothesized to be important in multidrug binding. A. WT-QacR-Et (orange) and WT-QacR-R6G (pink) complex structures illustrate how tyrosines move to accommodate different ligands. B. Since there are glutamates scattered throughout the binding pocket, ligands can interact with one or more regardless of the location of the ligand with the pocket.


Y123
Y103
A

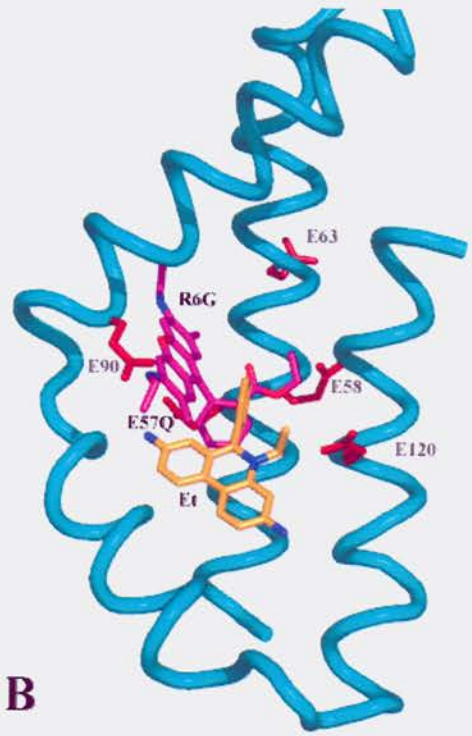

E63
R6G
E90
E58
E57Q
E120
Et
B

Figure 21. Conserved membrane-embedded acidic residues in transmembrane segment 1 (TM1) of drug-transport proteins. The primary sequence alignment of TM1 from MdfA and presumed TM1 of several drug transporters and putative drug transporters are shown. This alignment was performed manually and is not intended to reveal similarities other than the location of the negative charged residues within the putative TM segments. The conserved membrane-embedded acidic residues are shown in red. Listed accession numbers refer to the Swiss-Prot database.

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

Name		TM1	Accession number
MFS family			
MdfA	15	allfplclvly**e**fstyig	q46966
YjiO	11	tlffpmalily**d**faayls	p39386
CmlA	12	laatvlllspf**d**llaslg	p32482
TetA	6	plivilstval**d**avgigl	p02982
SgeI	12	islllt1flaal**d**ivivv	p33335
ToxA	53	mslgaaaflgal**d**atvva	148797
TetL	17	iwlcilsffsvln**e**mvln	u17153
Mmr	29	talatgfvmatl**d**vtvvn	p11545
TcmA	29	lavavgvmmval**d**stiva	p39886
Ptr	24	gviclaqltvll**d**ntvls	x84072
LfrA	23	avlalpvlliai**d**ntvla	u40487
QacA	22	vvlavslfvvtm**d**mtili	p23215
SmvA	9	viivlvyipvai**d**atvlh	p37594
CmcT	24	vlactakflvvf**d**tsvit	q04733
LmrA	34	vflavlqlliav**d**vtvvn	p46104
EmrB	17	ialslatfmqvl**d**stian	p27304
EmrY	21	ialslatfmqml**d**stisn	p52600
YieO	14	wlaamaffmqal**d**atiln	p31474
SMR family			
EmrE	4	yiylggaila**e**vigttlm	p23895
QacE	4	wlflviaivg**e**viatsal	u49101
Smr	3	yiyliiaist**e**vigsafl	p14319

Chapter 2: Introduction to Macromolecular Crystallography

After Max von Laue used X-ray diffraction to prove the ordered internal structure of crystals in 1912, X-ray crystallographers began to set their sights on the crystals of biological molecules—a pursuit that took several decades to truly progress. It was not until the 1930s that crystallographers first started crystallizing protein and viral particles. Then, in 1953, Rosalind Franklin's low-resolution X-ray data of the fiber diffraction of DNA was instrumental in cracking of the genetic code (*51-53*). But the atoms of the helix were not actually visualized with this data; the geometry of the spots identified only the structure as a helix and measured its dimensions.

Crystals allow better diffraction of X-rays and for much higher resolution data than the DNA fibers used in these experiments. Dr. John C. Kendrew illustrated this in 1957 when his crystals of sperm whale myoglobin diffracted beyond a resolution of 2 Å (*54*). These data were not only good enough to visualize α-helices for the first time but also good enough to actually identify most of the amino acids in the structure.

The advances in technology since then have dramatically accelerated the pace of crystallography. X-ray sources and detectors and computational power have improved so much that under the right conditions, a crystallographer may be able to solve a structure to a resolution of 2 Å in a matter of days or weeks rather than years. Moreover, the limits of resolution have been pushed back by the advent of powerful synchrotron radiation and state-of-the-art detectors. With an exceptionally good crystal, synchrotron source X-rays can produce data that yield such detailed information that even hydrogens can be distinguished. Using this new technology, scientists are studying some of the finer

details in structural biology, such as ligand recognition, which can require assessment of minute structural details.

This chapter introduces the concepts and technology intrinsic to the X-ray crystallographic techniques used in this dissertation. The first three sections provide explanations for important concepts assumed in the discussion of the later subsections. The introductory topics discussed include:

- Crystals of Biological Macromolecules
- Crystal Symmetry
- Principles of X-ray Diffraction

To give the readers a sense of the actual progression of events in crystallography, the following topics are arranged such that the sections mimic the chronological process of an actual experiment in macromolecular crystallography. The process of X-ray crystallography includes:

- Crystallization of Macromolecules
- X-ray Radiation and Detection
- Crystal Mounting
- Experimental Set-up
- Data collection and Processing
- Crystal Twinning
- Calculation of the Electron Density
- Determination of the Phase of the Electron Density Equation
- Model Refinement

Particular attention is paid to the concepts that may be helpful in understanding the full scope of the experimental section of the dissertation.

Crystals of Macromolecules

The keystone of the technique of macromolecular X-ray crystallography is the crystal itself. A crystal is a three-dimensional ordered array of molecules, the properties of which allow for the diffraction of X-rays required for structure determination at atomic resolution. Considering the complexity and flexibility inherent in most proteins, it is surprising that they can form such ordered structures. Furthermore, the contacts between protein molecules are generally spread far apart and are often mediated by water. Hence, protein crystals are much smaller and more fragile than other simple organic and inorganic crystals, such as sugar and salt respectively.

Crystal Symmetry

A crystal is a parallelepiped repeated in three dimensions. This prism whose faces are all parallelograms is called the unit cell, and it is the basic repeated unit that makes up the crystal (Figure 1A). This unit cell helps define the three dimensional grid (or lattice) of the crystal and determines many of the properties of its diffraction, as shall be discussed later. Within the unit cell may be further symmetry—that is, some parts of the unit cell may be symmetrically related to other parts. In these cases, the unit cell consists of a repeated portion that cannot be reduced into smaller symmetrical units (named appropriately, the *asymmetric unit*) (Figure 1B). The asymmetric unit consists of water; solutes such as molecules of salt, buffer, cryoprotectants etc., and, hopefully, the macromolecule of interest. The molecule of interest can be in an asymmetric unit once,

such as the dimer of QacR seen in one crystal form, or twice, such as the two dimers of QacR seen in another crystal form, or any number of times. In the crystal structure of EthR, another TetR family member, there is only 1/2 of a dimer in the asymmetric unit.

Crystals can be classified into seven three-dimensional coordinate systems: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, rhombohedral, and cubic. These are defined by six parameters of the unit cell (Figure 2). Three of the parameters are the lengths of the unit cell's edges, which are designated *a*, *b*, and *c*. The other three parameters are the interaxial angles. The angle α lies between *b* and *c*, β lies between *a* and *c*, and γ lies between *a* and *b*. The parameters are independent can have any value in the triclinic system, but the more symmetry inherent to a system, the more restrictions on the values of the parameters in the other crystal systems (Table 1). The parameters of a cell in the monoclinic system can be of any value, but two of the angles must be 90°. The parameters of an orthorhombic cell, however, are only free in its cell edge lengths. There are only two independent variables in tetragonal, trigonal, and rhombohedral systems. The hexagonal crystal system has set angles of 90° and 120°, but there are two independent values for the cell edges. In contrast, in the rhombohedral crystal system, the first independent variable is the length of the equilateral cell edges, and the second is the angles of the cells that are interdependent. The three angles in this system are equal and not equal to 90°. Otherwise this would be identical to the cubic system. For the cubic crystal system, as all the edges are the same length and the angles are fixed at 90°, there is only one independent variable: the length of any edge.

The unit cell is constrained to the geometric possibilities defined by the crystal systems, since deviating from them would result in a structure that cannot be

symmetrically repeated infinitely in all directions. Crystal systems, however, also have inherent symmetry defined by two simple types of operations, rotation and reflection. Rotation around an axis may be 2, 3, 4, or 6-fold symmetric, meaning the rotation of 360° may be divided into 2, 3, 4, or 6 parts. Reflection across a plane divides the unit cell. More complicated symmetry results from symmetry operations that are a combination of the two.

For each of the seven crystal systems, there is one primitive lattice—that is, one lattice point per cell. Non-primitive lattices are created by addition of extra lattice points to the unit cell. Though the non-primitive lattices are more complex, these lattices still conform to the symmetry of the seven crystal systems. A non-primitive lattice can be most simply viewed as "the combination of a primitive lattice with one or more offset identical copies of itself" (*55*). The sum of the seven primitive and seven non-primitive lattices yields the 14 Bravais lattices (Figure 3).

Combining all Bravais lattices with groups of symmetry operations called *point groups* leads to 230 unique arrangements of points in space, or *space groups*. The thirty-two point groups describe crystals in terms of symmetry operations relating their faces. The space groups defined by the combinations of point groups and Bravais lattices are the only ways an object can be arranged in an infinite lattice. Not all space groups are available to protein crystals due to the asymmetry of the protein itself. Interestingly, aspects of space groups indicate that structural components in the unit cell can be related by symmetry not seen in the external form of the crystal. For example, the screw axis, a symmetry operation that is often found in protein crystals, is a combination of rotation

and translation such that the symmetry element moves up the axis of rotation like a staircase. Yet the crystal faces only reflect the rotational component.

Principles of X-ray Diffraction and Data Collection

X-ray diffraction in crystallography is caused by the constructive interference of scattering from electrons of the molecules in the crystal. X-rays are absorbed by the electrons of the crystal and reemitted at the same frequency as the X-ray, hence the scattering is considered "elastic". The crystal acts as a three-dimensional lattice, so that the X-rays directed through the crystal diverge from the incident X-ray beam in a pattern that is based on the order and parameters of the crystal itself. However, the intensities of these X-rays give us critical information of the structures that comprise the crystal. In this portion of this chapter, we will consider what X-rays are and how they scatter from a unit cell and a whole crystal. We will also discuss how the X-rays scattered from the crystal form coherent X-rays from which the crystallographer can then derive the structure of the component molecules.

X-rays can be described many ways mathematically. One form of the equation is A cos $\alpha + i$A sin α where α is the phase of vector **A** and A cosα and A sinα are the real and imaginary components to the wave represented by vector **A**. This mathematical description of the wave can be easily depicted in an Argand diagram that illustrates phase (Figure 4). The wave equation can also be abbreviated to an exponential form that is a little more concise, A exp [$i\alpha$]. In describing the behavior of X-rays, the two forms will be used interchangeably.

The total scattering from the unit cell can be described as

$$\mathbf{F(S)} = \sum_{j=1}^{n} f_j \exp[2\pi i \mathbf{r}_j \cdot \mathbf{S}]$$

where j = atom # 1, 2, 3... n. and *f* the scattering from an atom (*56*). **S** is a vector defined by $\mathbf{s}_o - \mathbf{s}$, the vector defining the change in direction of incident and scattered X-ray vectors that retain their intensity after reflection (Figure 5). The vector $\mathbf{r}_j$ defines the position of the atom from the origin of the unit cell. The **F(S)** is the so-called *structure factor,* as it depends on the arrangement of atoms or structure of the unit cell. The structure factor is what the intensities of the X-rays determine, and from these measurements we calculate the electron density for the unit cell. But thus far, we see that all of the atoms of the unit cell contribute to scattering. This raises the question of how to measure diffraction, since the scattering is everywhere. It would be easier to measure the intensities of X-ray scattering if they were discrete intensities. Fortunately, the limitation of diffraction to discrete, measurable intensities is provided by diffraction from the whole crystal.

How diffraction by a crystal creates these discrete, measurable intensities that contain structural information about the molecules comprising the crystal can be seen from mathematical equations. Extrapolating scattering from a unit cell to that of a crystal lattice is simple, considering the ordered nature of the crystal itself. The crystal has translational vectors **a**, **b**, and **c** along the unit cell axis and the unit cell is repeated n_1 in the **a** direction, n_2 in the **b** direction, and n_3 in the **c** direction (Figure 6). Thus, for a unit cell with its own unit cell at position $t\cdot\mathbf{a} + u\cdot\mathbf{b} + v\cdot\mathbf{c}$, in which *t*, *u*, and *v* are whole numbers, the scattering is

$$\mathbf{K(S)} = \mathbf{F(S)} \times \sum_{t=0}^{n_1} \exp[2\pi i t \mathbf{a} \cdot \mathbf{S}] \times \sum_{u=0}^{n_2} \exp[2\pi i u \mathbf{b} \cdot \mathbf{S}] \times \sum_{v=0}^{n_3} \exp[2\pi i v \mathbf{c} \cdot \mathbf{S}] \ (56).$$

Since the number of unit cells in the crystal is very large, n_1, n_2, and n_3 are large, and the summation over **t**, **u**, or **v** is always equal to zero unless **a**·**S**, **b**·**S**, and **c**·**S** equal integers (designated, *h*, *k* and *l* respectively). These conditions for X-ray diffraction are known as the Laue conditions: **a**·**S** = *h*, **b**·**S** = *k*, and **c**·**S** = *l*. An Argand diagram demonstrates how, when the Laue conditions are not fulfilled, many equal-length vectors with different phases are lined head to tail they ultimately cancel each other out (Figure 7). Yet when the phases of these vectors are in phase they combine to make a much more intense X-ray. In summary, there is no net diffraction from the scattered X-rays, unless scattering off atoms in a plane with the indices *hkl*. In this case, the intensity of the total scattered wave is proportional to the number of unit cells in the crystal and the structure factor **F(S)**.

Though ultimately diffraction is from the electrons of all atoms in the crystal, the conditions that allow the constructive interference of the scattering of all of these X-rays to form a data point was best described by Lawrence Bragg as diffraction from planes. The conditions that allow for diffraction of the X-rays are summed up by the equation: $2d_{hkl}\sin\theta = n\lambda$. The interplanar distance is defined as d_{hkl}, the wavelength of the radiation is λ, and θ is the incident reflected angle of the radiation.

This equation can be explained through the simple geometry of the accompanying diagram (Figure 8). The paths defined by the incidence and reflection of X-rays 1 and 2 (R_1 and R_2) create the angles θ with their respective reflecting plane shown (P_1 and P_2). The ray R_1 reflects at point A on the upper plane and the R_2 reflects at point B. Ultimately, for these X-rays to reflect in phase, the path length added to the path of R_2 compared to R_1 must be an integral value of the wavelength of the radiation. This path

length can be determined geometrically by extending a line segment from R_2 to point A from point C, an added point on R_2. As the angle between reflection points A and B and P_2 is 90° and the angle between CB and P_2 is θ, it follows that angle ABC is 90-θ°. As part of the right triangle ABC, angles BCA and ABC add to 90°. Thus, angle BCA is equal to θ. It then must follow that the distance of AB/BC is equal to sinθ or distance BC is equal to the distance AB multiplied by sinθ. The distance AB is the same as d_{hkl} or the interplanar distance. The total increase to the path length of R_2 is then $2d_{hkl}\sin\theta$. To ensure the two rays are still in phase this length has to be an integral wavelength or nλ. Thus, $n\lambda = 2d_{hkl}\sin\theta$.

The significance of the reciprocal lattice can be seen from Bragg's law, where sinθ is proportional to $1/d$ and that given a constant λ. Since we know X-rays are diffracting from the lattice and since Bragg's law is true, then expressing the reciprocal lattice as $1/d$ should allow us to relate the paths of the actual X-rays to the lattice of the crystal. For this reason, crystallographers generally look at the lattice in reverse. An example of conversion from the real space to reciprocal space is easiest with an orthorhombic space group (Figure 9). The cell edge a* in the reciprocal lattice spacing becomes 1/a as is true for the other cell edges the angles remain 90°. These relationships become more complicated in other space groups, but well defined. In addition, these reciprocal space lattice points are equal to values of *hkl* of the Laue conditions, as well as the indices in the naming of planes.

Diffraction in terms of the reciprocal lattice can be useful to determine the path of the X-ray. In the description of the Ewald sphere (a 3D-corollary of Bragg's law), it should be apparent how the reciprocal space can aid us in predicting X-rays—that is, it

can how resultant X-rays correlate to the crystal lattice. We will look at a path of an X-ray in a crystal in reciprocal space that bisects a circle of radius $1/\lambda$ and passes through Points B and the Origin that lay on the circumference of the circle (Figure 10A). The origin is also the origin of the reciprocal lattice. P is a point on the semicircle that also corresponds with a reciprocal lattice point. The value of the angle OBP is represented by θ. Because angle OPB is inscribed in the semicircle it must be a right triangle. Therefore, sin OBP = $\sin\theta$ =OP/OB= OP/$(2/\lambda)$. As OP, the line from the origin to the reciprocal lattice point, is by definition $1/\lambda$, we get $2d_{hkl}\sin\theta=\lambda$ (or Bragg's law), we have the geometry for an X-ray reflection. Thus any time a reciprocal lattice point intersects the circle, there will be a reflection with a defined geometry. Figure 10B illustrates how to determine θ from the detector distance (DD) and the distance of the measured intensity (DC) from the center of the detector.

$$\tan(2\theta) = \frac{DC}{DD} \quad \text{or} \quad \theta = 2\cot\frac{DC}{DD}$$

Given the relationship between θ and d_{hkl} in Bragg's law, we can now calculate d_{hkl}. The measured intensities that are farthest from the center have the smallest d_{hkl}. This distance is the limiting resolution for the structure.

This is a two-dimensional example; of course, it is actually a sphere of reflection in three dimensions. As the crystal is turned in the X-rays, lattice points defined as having integral values of *hkl* will approach and recede from this sphere of reflection (Figure 11). In this way, spots will appear and disappear from the diffraction pattern, as the crystal turns through specific points in the reciprocal lattice. In real space, this describes from which planes the X-rays diffract and how different planes will reflect as the crystal is tuned in the X-ray beam.

In summary, X-rays are scattered by the electrons in the crystal. This scattering is only coherent from electrons in atoms lying in planes in the crystal that are consistent with Bragg's law. The coherent scattering forms discrete measurable intensities that are related to the electron density of the atoms in the reflecting plane. Through analysis of these data we can assemble a 3-dimentional structure of the protein into which a model of the protein can be constructed.

The Process of X-ray Crystallography

The following sections describe the process through which a crystallographer creates protein crystals, collects data, and analyses that data. Further theoretical discourse is scattered through the text when necessary.

Crystallization of Macromolecules

Crystallization of protein molecules will only occur in a supersaturated solution. In equilibrium between the solid and liquid phase of the protein, there would be no net increase in the solid phase—in this case the solid phase being crystalline. A saturated solution is in equilibrium between these two phases; thus one requirement of protein crystal formation will only occur when the protein in solution must surpass the point of saturation to become supersaturated.

When particular salts, polymers, and organic solvents are added to the solution in too great a quantity, they will cause protein to precipitate out of solution. These substances are therefore called *precipitating agents*. However, in controlled quantities, these agents are effective in forming crystals. In crystallography, we hope to induce a single crystal to form and grow, rather than to precipitate the protein. It is also important

to avoid creating many protein crystals, or *overnucleation*, as such crystals tend to be too small to collect data from. No matter method used, the crystallographer will have both of these outcomes. The goal, then, is to use techniques that can facilitate supersaturation without overnucleation and precipitation of the protein.

To nucleate a crystal, the solution must reach a higher protein saturation state than the one that is just above the protein limit of solubility in the solution. Just like water can become supersaturated before freezing, proteins can be supersaturated before crystallizing. In the case of proteins, the protein and precipitate concentrations are even higher than the one required for saturation, such that they reach the labile state or the state at which the protein will form (Figure 12). However, if the protein and precipitate concentrations are too high, protein overnucleation and precipitation will occur. The protein concentrations between the labile phase concentration and the saturation point concentration of the protein is known as the *metastable phase*. In this phase, crystals will continue to grow but no new crystals are formed. As growing one large crystal sufficient for reading data is generally the goal of a protein crystallographer, the ideal concentration of protein is represented on the phase diagram as just above the transition between the metastable and the labile phase. Thus, once the crystal starts growing and removes enough protein from the solution to reach the metastable phase, further crystals will not form, but the protein crystal will continue to grow until equilibrium between the solid and liquid state of the protein is reached.

Solubility of a protein can be affected in two ways. First, one can affect the availability of water to solvate the protein. Increasing the concentration of the protein beyond the point of saturation is the most obvious route to limiting the available water to

solvate protein, but it is also possible to lower the temperature and add hydrophilic solutes. The second method of changing protein solubility is by affecting the affinity of the protein for itself. The pH, ionic strength, and dielectric of the protein solution can be affected to maximize charge-charge attraction or minimize charge-charge repulsion of the charged amino acids of the surface of a protein.

Combining precipitants with controlled evaporation to create a supersaturated solution of protein is how this state of protein solubility at the edge of the labile phase is reached. This is the heart of two important crystallization techniques, hanging drop and sitting drop vapor diffusion (Figure 13). With these techniques, the water in protein solution competes with that of a concentrated precipitant in a buffer called the *mother liquor*. The protein solution is mixed with the precipitant and sealed in a chamber containing a solution of fully concentrated mother liquor at a much larger volume. The water molecules in the protein and mother liquor are constantly evaporating and condensing again into these solutions. The water molecules that evaporate out of either the protein or mother liquor are more likely to condense in the mother liquor, as it is more highly concentrated in solutes. Entropy increases more if the saltier solution is diluted. Thus, the water migrates from the protein solution to the mother liquor and gently concentrates the protein and precipitants to the edge of the labile phase. While these techniques are widely used, in practice they still waste a lot of protein on precipitates and showers of crystals. Every protein is different, so finding the right buffers and precipitants; the right concentration of the precipitants; and the right concentration of the protein requires trial and error to get a single large crystal.

X-ray Radiation and Detection

Bragg's law helps explain why nearly monochromatic X-rays are important for data collection. First of all, we specifically need X-rays for data collection because X-rays have very short wavelengths, and the lower the wavelength of the incident X-rays, the higher the potential resolution of the structure. If the wavelength is too long, Bragg's law shows us that the smallest interplanar spacing or resolution for the data will be much larger than the interatomic distances in the underlying structure. The wavelengths of X-rays are in the right range, as they are on the same scale as the interatomic distances in macromolecules. Another concept gleaned from Bragg's law is that X-rays used in X-ray crystallography have to be nearly monochromatic to get these discrete spots. Different X-ray sources solve this potential problem in different ways and the method of solving this problem is often dependent on the ways the X-rays are produced.

The conventional laboratory X-ray source produces X-rays by exciting inner K-shell orbital electrons of transition metals such as copper or molybdenum. After an arc of electricity removes an electron from the K-shell electrons from the L and M orbital fall to fill the vacancy. The change in potential energy from the higher to lower shell is offset by the emission of an X-ray. The X-rays produced by each orbital are nearly monochromatic, producing K_α with the L to K transition and K_β with the M to K transition (Figure 14A). There are two home X-ray sources that use this general method to produce X-rays; they are the rotating anode and sealed tube sources. The rotating anode source has about 5 times as much flux as the sealed tube. The former has the advantage of being able to use greater power by rotating the anode quickly to distribute and dissipate the heat produced from the electron bombardment.

The electron bombardment also creates a wider spectrum of lower intensity X-ray noise called *bremsstrahlung radiation.* The slowing of the electrons in the metal target causes this spectrum of radiation. K_α is the stronger peak of the X-rays, so ultimately we would like to remove the K_β and as much of the bremsstrahlung radiation as possible. Conveniently, a foil of a metal with the atomic number one less than that of the target metal in the X-ray production does well as a filter to make the X-rays nearly monochromatic (Figure 14B). The absorption edge of this element falls off significantly between K_α and K_β. Thus it absorbs much of K_β and the lower wavelength bremsstrahlung radiation, leaving K_α as the major population of X-rays emitted from the source (Figure 15). Crystal monochromators very nearly select completely monochromatic light, but there is a significant loss in X-ray intensity inherent to its use.

One source of X-rays that is intense enough to take a hit in intensity by a crystal monochromators is synchrotron radiation. Synchrotrons are large particle storage rings that spin off high intensity electro magnetic radiation. Charged particles emit electromagnetic radiation as they accelerate. A more familiar form of this may be the principle used in radio and television. The carrier radiation for these media is created by the acceleration of electrons back and forth on an antenna. To produce synchrotron radiation, an electron or positron traveling at relativistic speeds is accelerated through the curved path of a large ring. The wavelength of the resultant radiation is a function of the energy at which the synchrotron is operated and the radius of the ring. These rings provide a wide spectrum of radiation, but the peak intensity of the larger synchrotrons is in the range of 0.8 to 2.0Å, a range ideal for structural determination of macromolecules

To generate even more intense radiation, the electron is also often forced through insertion devices such as *wigglers* and *undulators*. Wigglers are a set of electromagnets that create a small detour for the electron by bending it sharply out of and then back into the ring. This extra acceleration created by the bend shifts the radiation to shorter wavelengths and increases the intensity of the X-rays in the wavelengths used by a crystallographer. Undulators are a little different from the wiggler, as they are series of small deviations from the path of the ring, such that the radiation created at each bend interferes with the radiation at the others. The resultant radiation is nearly monochromatic and the intensity is enhanced as much as 10,000 times.

The two methods of detecting X-rays general used at the synchrotron and the conventional X-ray crystallography laboratory are through the charge coupled device (CCD) detector and the imaging plate, respectively. However, the CCD detector is becoming much more common at the home lab, and imaging plates may still be found at some lines at the synchrotron. The three main components of the CCD detector are a phosphor screen that converts X-rays to light, a fiber-optic taper that shrinks the light image down to the size of the CCD chip, and a CCD chip that detects the light image as an electric charge image. The imaging plate is made of a material that changes energy states when irradiated with an X-ray, basically storing the energy proportional to that of the X-ray. The energy is stable until irradiated by a laser, at which time the data is read by scanning the plate. The major advantages of the CCD detector are its ability to read wider ranges of intensities and its very quick read out time. Of course, the CCD detector is much more expensive than the image plate detector.

Crystal Mounting

Between the detector and the X-rays and the detector is the goniometer, where the crystal is mounted. The goniometer allows adjustments to the position of the crystal in order to ensure that the crystal is properly centered for data collection and will not move out of the X-ray beam as the crystal is turned through the course of the experiment. For room temperature data collection, the crystal is drawn into a thin glass capillary that is then mounted vertically on the goniometer for data collection. But more commonly the crystals are mounted by supercooling of the crystal onto small loops.

As proteins are sensitive to oxygen radicals during long or high-energy X-ray exposures, cooling them to 100 K can help retard the process. However, ice crystal formation can destroy the crystal or create its own diffraction pattern; thus, the crystals are shielded from freezing by cryoprotectants. Usually, the crystal is soaked in a modified mother liquor with a low freezing component. For example, soaking a QacR crystal in a solution containing the same buffer and a slightly lower amount of the precipitating agent but with an added 20% glycerol by volume can protect the crystal from freezing. A protected crystal is scooped up with a loop on a small pin mounted to a magnetic base and either frozen directly on the goniometer, where a stream of cold air will keep the crystal at 100 K for the duration of the data collection, or it is flash frozen in liquid nitrogen first.

Experimental Set-up

After the crystal is mounted, a few test shots at a few distant angles are done to determine the course of the data collection. These shots will determine whether the crystal is worth collecting, the length of exposure for data collection, and the detector

distance that will ensure the most data are collected without overlap of diffraction spots. The test shots will also determine whether the crystal diffracts non-isomorphously and, if so, which orientation of the crystal would be best to collect. For isomorphously diffracting crystals, these test shots may also be analyzed to determine the most efficacious set of angles to collect through in order to achieve the fastest, most complete data collection. In higher symmetry crystals many of the reflections are symmetrically related so a complete data set requires a smaller set of angles. Data are collected by turning the crystal in the beam through the prescribed range of angles. While turning through this range of angles, images or *oscillations* are taken in small wedges, usually 0.5-1°.

When data collection is completed, computer programs integrate and scale the collected intensities into scaled intensities. This process works as follows. Initially, the spots are analyzed to determine the space group and unit cell dimensions. The programs will determine some of the fundamental aspects of the space group of the crystal. It cannot determine such symmetry elements as screws from one image but processing the data with lower symmetry elements will not affect the data. Later, with a fully collected data set, analysis of the symmetry and periodic absences of intensities yield the space group of the crystal. However, one only requires the unit cell dimensions determined by indexing the crystals if one is looking for an estimate for the number of molecules in the unit cell. The number of subunits (Z) in the unit cell is a function of the volume of the cell (V_{cell}), the molecular weight of the protein (M_r), and the density of the crystal (V_M).

$$Z = \frac{V_{cell}}{V_M \cdot M_r}$$ where V_M can be measured but usually close to 2.15 A^3/Da

After indexing, the *mosaicity*, or the degree of disorder to the crystal lattice, is estimated as a correction factor, then the mosaicity and the cell edges can be refined from a number of images. With the known mosaicity, cell edges and space group the computer can predict the location of all of the X-ray diffraction measurements collected with the detector. The recorded measurements (or spots) are then integrated to acquire the total intensities for each reflection.

It is during integration that the intensities are corrected both for polarization and for the *Lorenz factor*. The Lorenz factor is a correction for the different lengths of time it takes a reciprocal lattice point to cross through the sphere of reflection. The length of time depends on the position of the lattice point in reciprocal space and the direction in which it approaches the sphere. The reflections from the points that linger on the sphere of reflection will have higher intensities than the other reflections; thus the intensities must be corrected. We also need to correct for polarization because, X-rays are partly polarized by diffraction from crystals, since the parallel and perpendicular components of unpolarized light reflect differently from the crystal. The reflection of the component of the X-rays that have the electric vector parallel to the reflecting plane is reflected dependent only on the electron density and is independent of the reflection angle. The reflection of the perpendicular component is dependent on the electron density but also on $\cos^2 2\theta$ (*55*). As the two vectors are originally of equal intensity the maximum change in intensity from this effect is only two fold. The calculation of the polarization correction is dependent on the source of X-rays, crystal monochromators, and the inherent polarized nature of synchrotron X-rays.

Scaling and post refinement helps fix the inconsistencies between the images in data collection. This process fixes variability beam intensity and diffracting power of the crystal to ensure that the absolute intensities are more consistent from one image to the next. An obvious way of obtaining this consistency is by comparing reflections of the same index that are found on a number of images and then rescale intensities so that reflections of the same index have the same intensity. Following scaling is post-refinement, which recovers partially recorded reflections, or *partials* to create a value for a completely recorded reflection.

After scaling and post-refinement the crystallographer has a list of structure factors amplitudes ($|F_{hkl}|$) and associated errors (σ) for every reciprocal lattice point measured. The error of data collection is calculated during the scaling process and is expressed as R_{sym}, which compares all the intensities of symmetry related reflections.

$$R_{sym} = \frac{\sum \left| \langle I(h) \rangle - I(h) \right|}{\sum \langle I(h) \rangle}$$

where $\langle I(h) \rangle$ is the mean of symmetry-related reflections and the h summation taken over all measured reflections. Generally, data collection is successful if the average R_{sym} is under 10%. The requirements for the data at the highest resolution are that the R_{sym} is under 50% and the intensity of the reflections divided by the error of the reflections (I/σ) is near 2.

Crystal Twinning

After scaling the data from higher resolution space group crystals, it is important to test for crystal *twinning*, which is the state of having two or more domains with differently oriented lattices in the same crystal. Twinning has the potential to render data

from a crystal useless. There are few types of twinning—epitaxial, pseudo-merohedral and merohedral—but merohedral twinning is the worst form of twinning. It is often not recognized and it can lead to more fundamental crystallographic problems. In merohedral twinning, the lattices of the different domains in the crystal align exactly so the diffraction pattern looks normal but contains contributions from crystallographically unrelated reflections (Figure 16). This often goes undetected, so that it is impossible to figure out why the structure cannot be solved (*57*). Sometimes the structure will appear to have been solved but then fail to refine to appropriate statistics. Nevertheless, structures have been solved with twinned crystals without correction, and with lower amounts of twinning there is little apparent effect on the quality of the data. In some cases, the twinning reflections can be resolved and the data treated normally, but often with higher degrees of twinning it is better to find untwinned crystals.

Merohedral twinning in crystallographic data may be detected in a number of ways. The simplest and most reliable methods are those that compare the intensities or structure factors of the twinned and untwinned data. The observed intensities in twinned crystals are the sums of two different intensities and do not obey ordinary Wilson statistics, which are part of the output of the scaling program SCALA (from CCP4). Twinning is apparent from the cumulative intensity distribution that plots the percentage of acentric and centric reflections against the normalized intensity (Figure 17). With twinning present, the plot of the distribution of the acentric data is sigmoidally shaped. The following equations, when applied to the acentric data can diagnose for twinning:

$\langle I^2 \rangle / \langle I \rangle^2 =$ 2 and 1.5 for untwinned and perfectly twinned data respectively

$\langle F^2 \rangle / \langle F \rangle^2$ =0.785 and 0.885 for untwinned and perfectly twinned data respectively (*58*).

Luckily, the only reported type of merohedral twinning in protein crystals is hemihedral twinning, in which there are only two differently oriented crystal domains. The fractional volume of the domain in the second orientation (or orientation with the smaller volume) is represented by α, the value of which, by definition ranges, from 0 to 0.5. Therefore "partial twinning" corresponds to twinning of $\alpha > 0$ to < 0.5. When α is 0.5, the crystal is considered to be perfectly twinned and is probably the most difficult twinning situation and in this situation people will most often look for new crystals. In cases of lower twinning fractions ($\alpha < 0.3$-0.4), the intensities from the two domains may be deconvoluted, as was the case with the crystals of the E120Q-QacR-MG complex described in this thesis.

The process of detwinning data starts with estimating the twinning fraction. If $I_{obs}(\mathbf{h}_1)$ and I_{obs} $(\mathbf{h}_2)$ are the twin related acentric observed intensities, H can be defined for each pair of reflections such that

$$H=|I_{obs}(\mathbf{h}_1) - I_{obs}(\mathbf{h}_2)|/(I_{obs}(\mathbf{h}_1) + I_{obs}(\mathbf{h}_2)) \quad (58)$$

H ranges from 0 to (1-2α) in the absence of errors. As H is a function of α and the true crystallographic intensities, the expected cumulative distribution is as follows: $S(H)=H/(1-2\alpha)$. To estimate the twinning fraction, one can use observed data and a possible symmetry relationship between the crystal domains to calculate H easily and compare it to standard curve. One can also use the expected probability distribution of H to determine the average value of H or the averaged squared H in terms of α. This allows α to be estimated directly from the equations

$\langle H \rangle = 1/2 - \alpha$ and $\langle H \rangle = (1-2\alpha)^2/3$ (*58*).

Once the twinning fraction is estimated accurately and not close to 0.5, the following linear equations can be applied to recover the true measured values:

$I(\mathbf{h}_1)=[1-\alpha)I_{obs}(\mathbf{h}_1)-\alpha I_{obs}(\mathbf{h}_2)]/(1-2\alpha)$ and

$I(\mathbf{h}_2)=[-\alpha I_{obs}(\mathbf{h}_1)+(1-\alpha)I_{obs}(\mathbf{h}_2)]/(1-2\alpha)$ (*58*).

The other types of twinning are usually obvious from the outset and generally observed as splitting of spots in the diffraction pattern or a data set that is very difficult to index. Epitaxial twinning has multiple domains that do not overlap or only partially overlap and pseudo-merohedrally related domains overlap almost completely but not quite. With either of these types of twinning, a crystal may have reflections from one domain dominate the diffraction so it will index and process without any difficulty. In this work a number of crystals of the E90Q-QacR-Et complex were collected until one domain dominated in the diffraction pattern and the data could be indexed and collected.

Calculation of the Electron Density and Phase Determination

Proper data processing of the structure factor amplitudes for every measured reflection (*hkl*) brings the crystallographer close to mapping the electron density. But it is still necessary to determine the phases of the measured structure factor amplitudes. Structure factor amplitudes contain no phase information, yet the structure factors in the electron density equation inherently contain such information:

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

The electron density equation can be rewritten to separate structure factor amplitudes ($|F|$) and the phases for each (α').

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

This electron density equation is a Fourier series. It is a sum of the waves that we now know as structure factors. As a Fourier series, the equation can be converted by a Fourier transform to an equation that measures structure factors as a function of coordinates and electron density.

$$F_{hkl} = \int_h \int_k \int_l \rho(x,y,z) e^{2\pi i(hx+ky+lz)} dxdydz$$

In phasing and model refinement the Fourier relationship between these two equations is critical as it important to be able to move back and forth between the two equations.

The primary techniques for phasing the structure factors in macromolecular crystallography are multiple isomorphous replacement (MIR), multi-wavelength anomalous diffraction (MAD), and molecular replacement (MR). In MIR, heavy atom additives containing mercury, gold, lead and platinum derivatives are commonly added to the crystal with the hope that the heavy atoms bind to specific sites in the unit cell. MIR uses the disparity between the powerful X-ray scattering ability of these electron rich atoms with the lighter scattering of the normal constituents of a protein (C, O, N, S) to gain phase information about the system. To be effective, the derivatives cannot affect the conformation of the protein or change the unit cell by more than 1/4 of the maximum resolution; hence the derivative must be isomorphous with the native crystal.

The structure factors of the protein with the heavy atom (F_{HP}) are equal to the structure factors of the heavy atoms (F_H) plus the structure factors of the protein (F_P). Using the native and heavy-atom derivitized crystals, we can measure the structure factor amplitudes for the protein ($|F_P|$) and for the protein with heavy atoms ($|F_{HP}|$). We can

then use these measurements to locate the heavy atoms with a *difference Patterson* map (the Patterson function will be discussed shortly). Given the locations of the heavy atoms and the known electron density for any atom, the amplitude and phase of their scattering can be calculated by a Fourier transform, we can calculate F_H. This in turn can help isolate the phase of the protein, as can be visualized in the following (Figure 18). Thus, if we plot the potential phase space of the protein atom as a circle with a radius $|F_P|$ from the center, we can plot a vector for $-F_H$, which has the length of F_H but points in the opposite direction (Figure 18A). From the end of this vector, we can create another circle with a radius of $|F_{HP}|$. This circle intersects with $|F_P|$ in two places, thus giving two possible phase solutions for F_P. Therefore another derivative is needed to decide which of these two phases is correct; hence, multiple isomorphous replacement (Figure 18B).

The Patterson function is a variation on the Fourier series with the form:

$$P(uvw) = \frac{1}{V}\sum_h\sum_k\sum_l \left|F_{hkl}^{\ 2}\right| e^{-2\pi i(hu+kv+lw)}$$

u, v, and w are positions in Patterson space. Patterson space is different from xyz space because it is related to the interatomic vectors rather than distances from an origin. With all of the interatomic vectors of a protein, this landscape may be complex, but with a difference Patterson between F_{HP} and F_P, the heavy atoms are easily visible.

$$P(u,v,w) = \frac{1}{V}\sum_h\sum_k\sum_l (\left|F_{HP} - F_P\right|)^2 e^{-2\pi i(hu+kv+lw)}$$

This difference Patterson removes all of the contributions to the map by the protein thus leaving only the heavy atom components. The positions of the heavy atoms are determined from the few heavy atom peaks seen in the map.

MAD can be considered a special case of MIR. Instead of relying on the introduced heavy atoms ability to scatter X-rays well, this method relies on an element's ability to scatter X-rays anomalously at wavelengths close to the absorption edge of the element. At wavelengths close to this edge, inner electrons of the heavy atoms will absorb the X-rays at a specific wavelength and emit them with a different phase than the incident X-rays. Thus, one can take the same protein crystal with a heavy atom derivative and expose it a few times with two to four different wavelengths of X-rays to get different amounts of anomalous scattering and use this data to find the heavy atoms. In this way, "MAD can be considered an ideal case of MIR: there is perfect structural isomorphism because only the X-ray wavelength is changed and the structure of the crystal is not perturbed" (*59*). Another advantage of MAD is that only one derivative is necessary.

We can therefore consider the data from exposures of the same crystal to different wavelengths of radiation as different derivatives that can be used to find the heavy atoms. Often, three wavelengths are chosen to get the maximum number of "derivatives," though, as we will see, often only two wavelengths are necessary. The remote wavelength is far from the absorption wavelength of the anomalous scattering atom, and structure factor amplitudes measured at this wavelength may be considered the native structure factors. Then the absorption peak and the absorption edge inflection points are taken, as they have different anomalous scattering characteristics from each other. The difference in the vectors between the structure factor amplitudes measured with anomalous scattering and those amplitudes measured at the remote wavelength are $\Delta\mathbf{F}'_{hkl}+\Delta\mathbf{F}''_{hkl}$. Together, these vectors can be considered similar to the heavy atom vectors in MIR such that, $|F_{hkl}|=|F_{hkl}^{anom}|-(\Delta\mathbf{F}'_{hkl}+\Delta\mathbf{F}''_{hkl})$(*55; 59*). Realizing that $|F_{hkl}| \neq |F_{-h-k-l}|$ in the case of anomalous scattering from noncentrosymmetric structures (such as proteins), F_{hkl} and F_{-h-k-l}

at the anomalous scattering wavelengths can be considered different derivatives. Thus, with only the native set and one anomalous scattering wavelength, there are now two different derivatives, which may enough to unambiguously solve the phase for the structure factors.

Due to the great number of structures already solved, molecular replacement is a common method for determining the phases for protein structures. Basically, one can use the phases from the structure of a known protein to get the phases for the structure factors of a protein with a similar structure. A simplified version of this would be phase replacement, where the phases of the model are used for the structure factors directly. This is commonly done for mutants of a protein where the crystals of the two proteins are isomorphous.

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F_{hkl}^{\text{obs}}\right|e^{-2\pi i(hx+ky+lz-\alpha'^{\text{model}}_{hkl})}$$

Molecular replacement is similar but in this case the computer searches through rotational and translational space to best match the model to the observed data. In order to be successful, the model is has to be structurally homologous to at least a part (usually a large part) of the molecule in the crystal. Though MR is a computational problem and does not require getting phases through extra crystallographic experiments, it is still complicated. There are 6 dimensions of rotational and translational space to consider. Moreover, like MIR and MAD Patterson calculations of the vectors of the protein and model are calculated, and all of the inter- and intramolecular distances complicate the maps. Thus, the components of MR are divided into easier computational problems yielding more easily interpretable information. The rotational searching and translational searching are divided into two separate events. Furthermore the data used in the rotational function are truncated in order to lower the amount of intermolecular vectors calculated by the function.

The rotation function is based on a *self-Patterson* or a Patterson function that involves the intramoleclular vectors of the protein. Because there are no intermolecular vectors this Patterson map would be the same for the same molecule in different crystal

structures, apart from the rotational difference. To search for that rotational difference, we use a rotation function derived from a self-Patterson based on the understanding that the maximum overlap between two self-Pattersons will occur when they have the same orientation. Thus an overlap function R of $P(\mathbf{u})$ [abbreviated from $P(uvw)$] with the rotated version $P_r(\mathbf{u})$ is defined as

$$R(\alpha,\beta,\gamma) = \int_U P(\mathbf{u}) \times P_r(\mathbf{u}_r) d\mathbf{u}$$

where U is defined as the volume of the Patterson map where the self-Patterson vectors are located (*56*). By assuming that all of the electron density outside of the volume $U = 0$ and expressing these Patterson functions in terms of structure factors one gets the following equation (*56*).

$$R(\alpha,\beta,\gamma) = \frac{U}{V^3} \sum_{\mathbf{h}} \sum_{\mathbf{h'}} |F(\mathbf{h})|^2 |F[C]\mathbf{h'})|^2 \times \mathbf{G}[-\mathbf{h} + \mathbf{h'})]$$

The rotation matrix, [C], that is applied to the Patterson function gives $P_r(\mathbf{u}_r)$ by definition. Therefore, [C] is the variable optimized in this equation. The individual reflections, *hkl*, are represented by **h** and **h'** for the data and the search model respectively. **G** represents the electron density of the volume U. Placing **G** in the equation as a factor of **–h** and **h'** ensures that the rotation function lands on a reciprocal lattice point and only includes the volume of U in order to make the function a self-Patterson.

There are several conventions for all of the aspects of the rotation angles. The most commonly used set of conventions is the Eulerian angles, because with this convention, the symmetry of the rotation function appears clearly. The system used by Rossman and Blow applies first a rotation around axis z by angle α, then progresses around the new position of axis-x by angle β then back again around z by angle γ, some call this the *zxz* convention (*56*). Others have used a similar convention, known as the *zyz* convention, which uses the z and y-axes in a similar manner. For the most part, however,

this rotation function has been superceded by the more powerful fast rotation function. Most programs use this function, including MOLREP, the program used in this thesis. This function basically expands the Patterson functions as spherical harmonics, thus three orthogonal rotational angles can be sampled simultaneously (*60*). This greatly increases the speed of rotational sampling.

Once the solution for the rotation function has been produced then the translation search proceeds. The translation function is basically a trial and error process where the model is moved and the structure factor amplitudes of the search model in every new position is compared to the amplitudes of the observed data. An alternative to the trial and error method is the translation function that correlates Patterson vectors of the search model with the Patterson vectors of the observed structure factors. Crystallographic symmetry is required for the translation function to work as the Patterson vectors are derived from molecules in the model related by symmetry. Yet, this is not a problem for the asymmetric triclinic system. As there is no symmetry in the triclinic crystal system, the origin may be chosen anywhere for the P_1 space group and the calculation of the translation function is not necessary. In higher symmetry space groups with greater than two subunits in the asymmetric unit the, relationships between the molecule in each symmetric unit may be determined in a pair wise fashion.

The first step is to create a *cross-Patterson* function that corresponds to the intermolecular vectors of symmetrically related search models. In other words, we would like the Patterson map of the vectors from the electron density of the model of the symmetry related copy (ρ_2) to the electron density of the original model (ρ_1) with out the intramolecular vectors of each model (Figure 19).

$$P_{2\to1}(\mathbf{u}) = \int_{cell} \rho_1(\mathbf{x})\rho_2(\mathbf{x}-\mathbf{u})d\mathbf{x}$$
$$= \frac{1}{V}\sum_{\mathbf{h}} \mathbf{F}_1(\mathbf{h})\mathbf{F}_2^*(\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{u})$$

where **u** is the vector in Patterson space (*uvw*) between the symmetrically related models and **x** is the location of the original model as related to the symmetry element.

The second step is to use this function to create a function that calculates the required translation of the model, relative to the symmetry axis, to fit the observed data. When the correct translation is found there should be a large peak because the vector sets from the model and the data will coincide.

$$T(\mathbf{t}) = \int_{cell} \mathbf{P}_{2\to1}(\mathbf{u}-\mathbf{t})P_{obs}(\mathbf{u})d\mathbf{u}$$
$$= \frac{1}{V}\sum_{\mathbf{h}} F_1^*(\mathbf{h})F_2(\mathbf{h})|F_O(\mathbf{h})|^2\exp(-2\pi\mathbf{h}\cdot\mathbf{t})$$

where **t** is the translational vector between the observed and calculated Patterson maps.

The two popular methods that indicate how well the model fits the data are the *R* factor and the correlation coefficient.

$$R = \frac{\sum_{hkl}\left||F_{obs}| - k|F_{calc}|\right|}{\sum_{hkl}|F_{obs}|} \qquad C = \frac{\sum_{hkl}(|F_{obs}|^2 - \overline{|F_{obs}|^2}) \times (|F_{calc}|^2 - \overline{|F_{calc}|^2})}{\left[\sum_{hkl}(|F_{obs}|^2 - \overline{|F_{obs}|^2})^2 \sum_{hkl}(|F_{calc}|^2 - \overline{|F_{calc}|^2})^2\right]^{1/2}}$$

The R factor indicates better fitting of the model when it is minimized. When a MR solution is found the R-factor should be below 0.59 as this is the value for a completely random solution. A successful solution will usually fall in the range of 0.3-0.4 though solutions higher may still work. The correlation coefficient is maximized rather than minimized like the *R*-factor. The correlation coefficient is used in MOLREP and has an advantage over the *R*-factor in that it is insensitive to scaling.

MOLREP also includes a translation function called the *packing function*. This function works well with the other translation functions in the program, as it can eliminate the need to search particular portions of translational space (*56*). This function basically tests the translational space of the protein in the unit cell to look for positions at which the intermolecular encroachment is at a minimum. This is a powerful aid for the translation function in higher symmetry space groups.

With computing power becoming greater and greater, brute force MR programs that sample both rotational and translational space simultaneously are likely to be more widely used in the near future. However, currently there is an interesting alternative to these programs. There is program based on evolutionary approaches to sampling the rotational and translational space simultaneously, called EPMR (evolutionary programming for molecular replacement) (*61*). This program creates a set of models in the crystallographic unit cell that evenly sample rotational and translational space for the correct MR solution. Test sets are randomly selected from these models and are compared to the data with the correlation coefficient, and the best is selected and the rest eliminated. The model population is held constant by adding back copies of the winning model with slight modifications to the rotational and translational parameters. After a specified number of generations, the model with the best fit is chosen and refined with conjugate-gradient optimization (*61*).

The MR model may still be fairly different from the actual position for the electron density. It is important to perform rigid body refinement to move the model into better register with the electron density. This refinement can also be used to fix small changes in the protein conformation. For example, I have two slightly different

conformations of QacR mutants. I found that sometimes, it was necessary to break the protein in areas of flexibility to allow the two fragments of the protein to refine separately. Thus the model could be adjusted for changes in conformation before finer refinement. One caveat is that this may separate atoms at the break to distances that will be necessary to repair before further refinement.

Electron Density Maps

The electron density calculated from structure factor amplitudes and the phases that were determined can be represented in a number of ways. If good unbiased phases are generated from MAD or MIR phasing techniques, one can represent the electron density of the structure as a F_o map where the electron density is calculated solely from the observed structure factors. However, with poorer phasing power the phases will need to be refined based on the model as it is added into the map. Furthermore, maps of structures phased by molecular replacement are inherently biased. Thus in these latter cases it is important to generate electron density maps that try to minimize model bias.

These maps are generated by various difference Fourier syntheses. In the amplitude portion of the Fourier series the amplitude of the observed structure factors is subtracted from the amplitude of the calculated structure factors. This reduces the model bias from the maps. Of course the phases are still determined by the model so model bias is never eliminated.

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|n|F_o| - |F_c|\right|e^{-2\pi i(hx+ky+lz-\alpha'^{\text{model}}_{hkl})}$$

where n is an integer. When n is one, the resultant F_o-F_c electron density map indicates the portions of the electron density into which the model has not been built. When n is 2,

the resultant $2F_o$-F_c electron density map represents the total electron density of the structure. The $2F_o$-F_c map is useful to work with as it is usually quick to calculate but it is still model biased, and this bias can interfere with model building. If there is a questionable area that might bias the data one can eliminate it and calculate an *omit map*. A similar map can be created that tries to remove bias from all portions of the model by creating a map that is a combination of a series of omit maps. This is still model biased but it is the least biased of the maps where the model is used for calculating phases.

Model Refinement

Different refinement programs perform further refinement of the protein differently but they all have common features. First of all, the programs use restraints to offset the large numbers of parameters that need to be refined in a protein. These parameters are the x, y, and z coordinates and the thermal factor, or B-factor, for each atom. Multiplying by the number of atoms in the asymmetric unit results in thousands of parameters that need to be defined by the data that was collected. A rule of thumb is that must be about three reflections for each parameter. At higher resolution data, the number of reflections is so great that more parameters can be added such as separate refinements for different vectors in the thermal parameter. However, with lower resolution data, parameters may need to be removed by refining B-factors in large groups rather than individually per atom. Another way in which to increase the ratio of data to parameters is to exploit the chemical information already known about proteins (*62*). Bonds between different atoms have very specific bond lengths without much variation, and similarly the van der Waals radius for any particular atom is constant. Amino acids have

stereochemistry that must be maintained through refinement (*63*). The restraints that use the above information limit the amount of space the refinement program can search.

Many structure refinement programs, such as TNT, used these restraints in conjunction with conjugate gradient refinement based on minimization of least-squares residuals of structure factor amplitudes to refine the data.

$$E^{LSQ} = E_{\text{restraints}} + w_a \sum_{hkl} (|F_o| - k|F_c|)^2$$

But this function may decrease as a result of systematic errors added to the model that better fit the data but that worsen the model (*64*). These errors can be accounted for by the fact that least-squares residuals do not consider the phase-errors in the calculated structure factors (*65*). Thus, least-squares residual refinement does not work well when the model is distant from the correct answer (*65*). The radius of convergence, or the maximum distance of the starting model to be able to find the answer, for least squares fitting is $d_{min}/4$.

Maximum-likelihood-based refinement combined with simulated annealing, which is included in the refinement program CNS, significantly enhances refinement over the previous least-squares methods. The maximum likelihood target significantly increases the radius of convergence, in effect raising the local minima in comparison to the global minimum (picture) (*66*). Furthermore, simulated annealing samples more refinement space than does the conjugate gradient alone. This removes the necessity to cut the data to lower resolutions for initial refinement in order to increase the radius of convergence (*66*). The cross-validation component, or R_{free}, was added to the maximum-likelihood minimization function to detect over-fitting of the model to the data and avoids

the systematic addition of errors to improve the fit. Cross-validation is achieved by testing a set of diffraction data that were never included in the refinement process.

Maximum likelihood measures the goodness of fit of a model to the data in a statistical way. This method uses the model to determine the probability of making a set of measurements given estimates of the model's errors and the errors of the measured intensities. The effects of the model errors on the calculated structure factors are made with σ_A, which includes cross-validation in its computation. The σ_A values are used to compute the expected value of $\langle |F_o| \rangle^{cv}$ and the corresponding variance ($\sigma^2_{ML_{cv}}$)(*65*). The form of the maximum-likelihood equation is a similar that of the least squares. In fact, least squares can be considered a special case of maximum likelihood.

$$E^{ML} = E_{\text{restraints}} + w_a \sum_{hkl \in workingset} \left(\frac{1}{\sigma^2_{ML^{cv}}} \right) (|F_o| - \langle |F_c| \rangle^{cv})^2$$

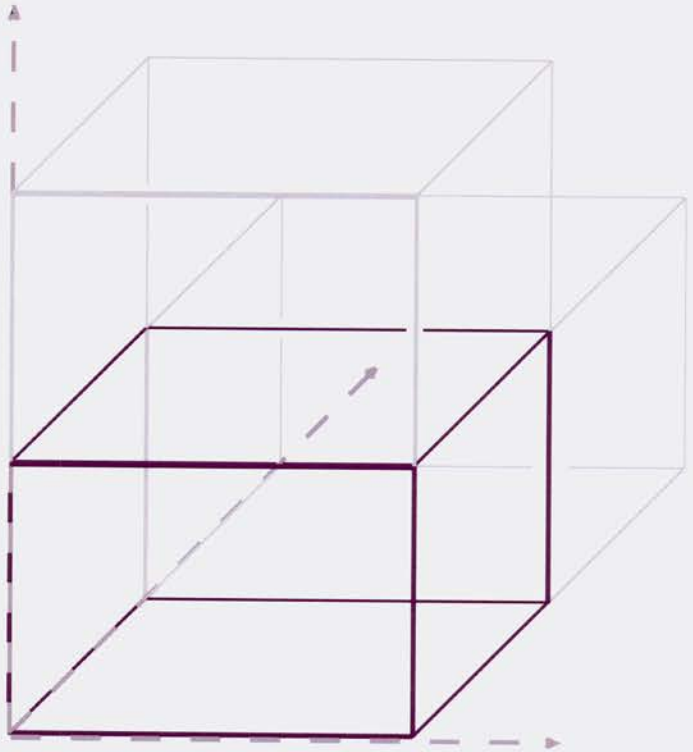
where w_a is the coefficient that weights the geometric components against the X-ray data components in refinement and $\sigma^2_{ML^{cv}}$ is the variance corresponding with $\langle |F_o| \rangle^{cv}$.

Structure Validation

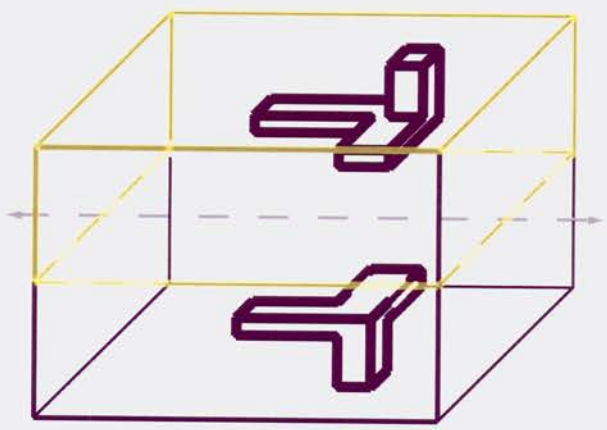
At the end of refinement, the structure is analyzed with structure validation tools before publishing. These tools, such as PROCHECK from CCP4 and the modelstats script in CNS, check for deviations from the normal statistics for bond lengths, interatomic distances, dihedral angles, etc (*67; 68*). Though the protein may just deviate from the norm, these data allow the crystallographer to check the model in specific areas before publishing.

The following chapter describes the methods used in this thesis including the conditions that were used to grow the protein crystals used in this thesis. Other methods include the techniques used to measure protein-ligand affinities as well as the protein purification protocol.

Figure 1. Unit cell and asymmetric unit. A. The unit cell is the basic building block of the crystal. B. The asymmetric unit (yellow) in a unit cell with a two fold axis.



A



B

Figure 2. Unit cell parameters.

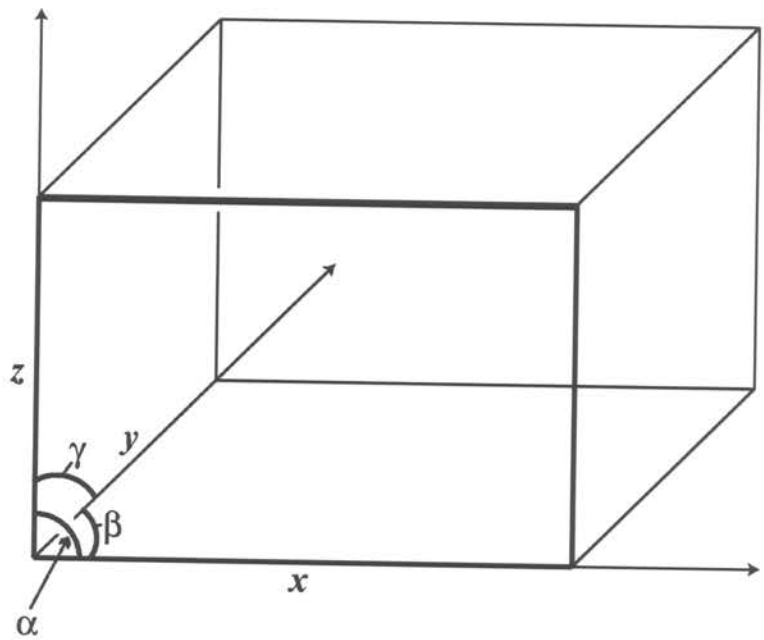

z
γ
y
β
x
α

Figure 3. 14 Bravais lattices.

Adapted from Stout, G. H., and Jensen, L. H. (1989) X-ray Structure Determination: A Practical Guide, 2nd Edition.

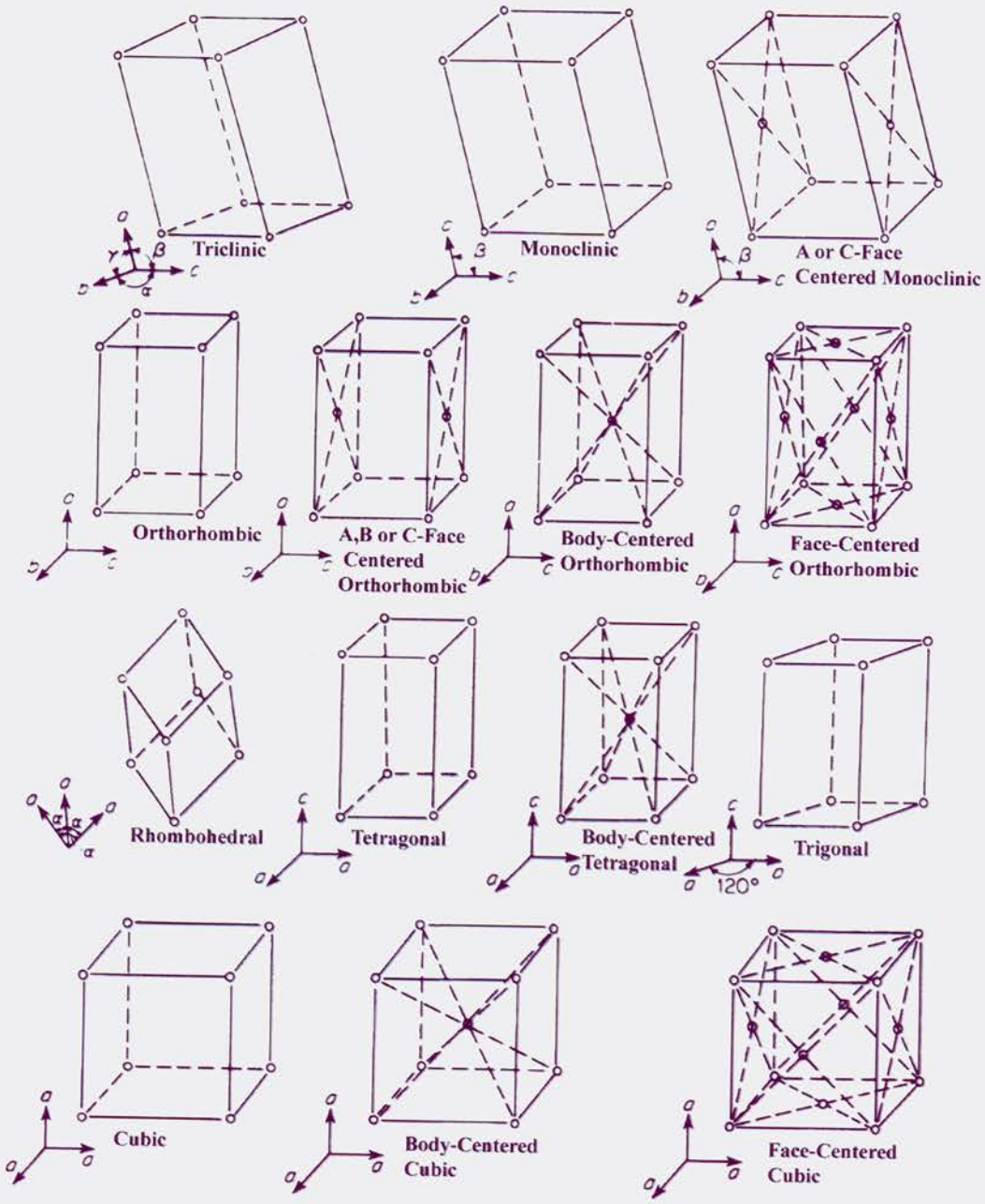

Triclinic
Monoclinic
A or C-Face
Centered Monoclinic
Orthorhombic
A,B or C-Face
Centered
Orthorhombic
Body-Centered
Orthorhombic
Face-Centered
Orthorhombic
Rhombohedral
Tetragonal
Body-Centered
Tetragonal
Trigonal
120°
Cubic
Body-Centered
Cubic
Face-Centered
Cubic

Figure 4. Argand diagram. One formula for a wave is A cos α + iA sin, α where α is the phase of vector **A.**

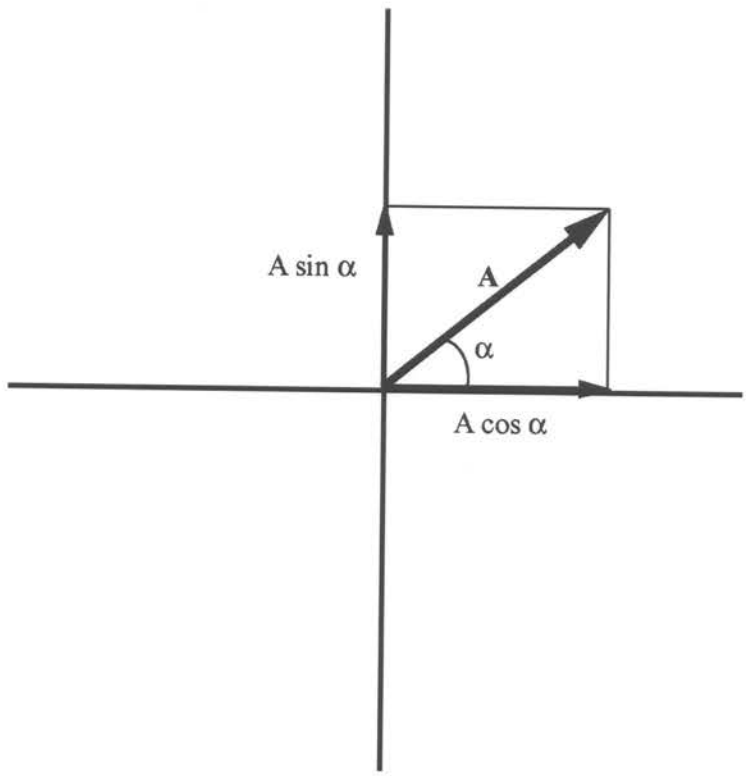

A sin α
A
α
A cos α

Figure 5. Vector diagram of **S**. **S** is a vector defined by $\mathbf{s}_o$- **s**, the vector defining the change in direction of incident and scattered X-ray vectors that retain their velocity after reflection

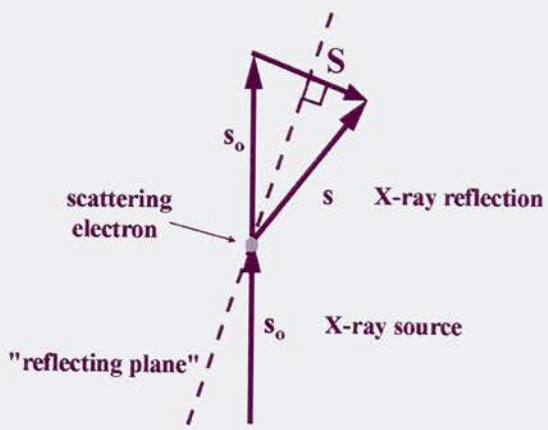

S
s_o
s
X-ray reflection
scattering
electron
s_o
X-ray source
"reflecting plane"

Figure 6. Vector description of positions of unit cells in the crystal. Unit cells stretch from 0-n_1, 0-n_2, and 0-n_3 along the **a**, **b**, and **c** vectors, respectively. These ranges represent the indices *t*, *u*, and *v*.

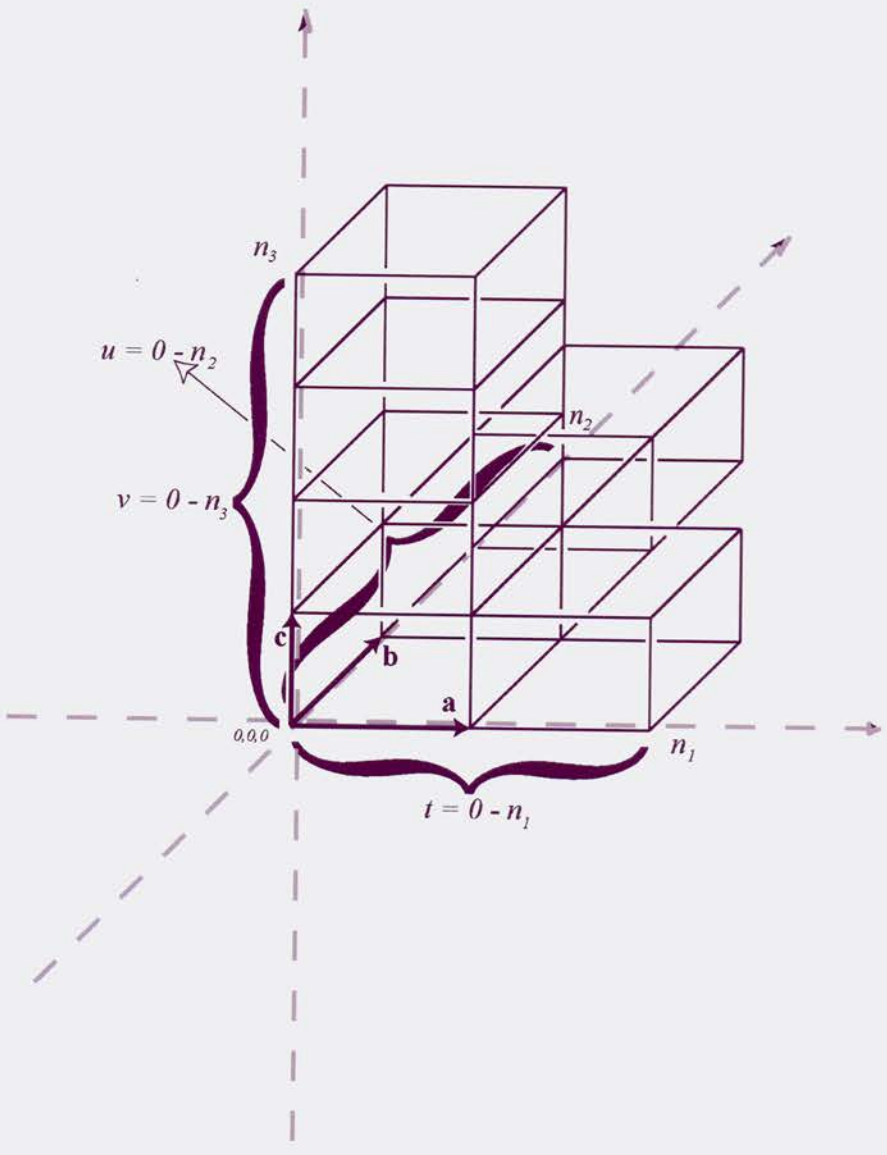
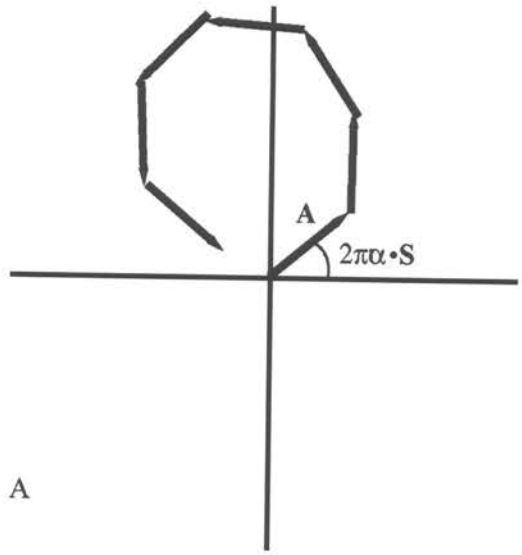
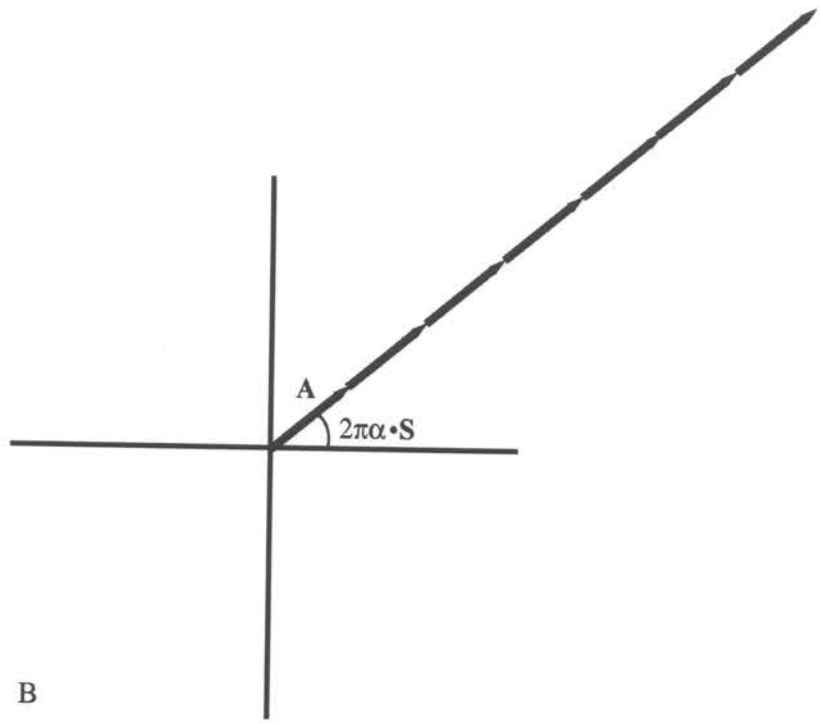

n_3
$u = 0 - n_2$
n_2
$v = 0 - n_3$
c
b
a
0,0,0
n_1
$t = 0 - n_1$

Figure 7. Vector addition and cancellation. A. When scattered waves from successive unit cells are in phase, their amplitudes are additive. B. When scattered waves from successive unit cells are not in phase, the waves cancel.



A



B

Figure 8. Bragg's law diagram. The dotted lines represent a lattice plane. Black dots represent individual atoms in the plane. The thin solid black lines represent the path of parallel X-rays.

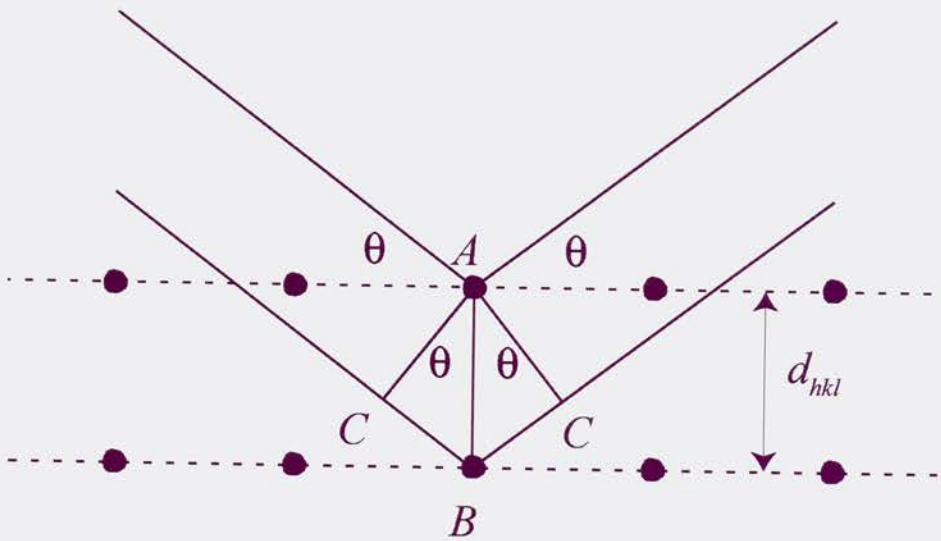
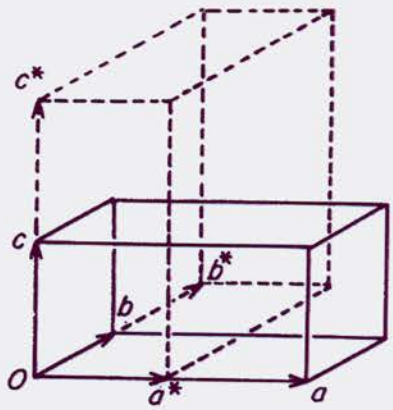

θ
A
θ
θ
θ
C
C
B
d_{hkl}

Figure 9. The relationship between real and reciprocal space. Examples of the relationship between real and reciprocal space in the orthorhombic (top) and monoclinic (bottom) space groups. Solid lines represent the unit cell in real space and the dotted lines represent the unit cell in reciprocal space


c*
c
b*
b
O
a*
a

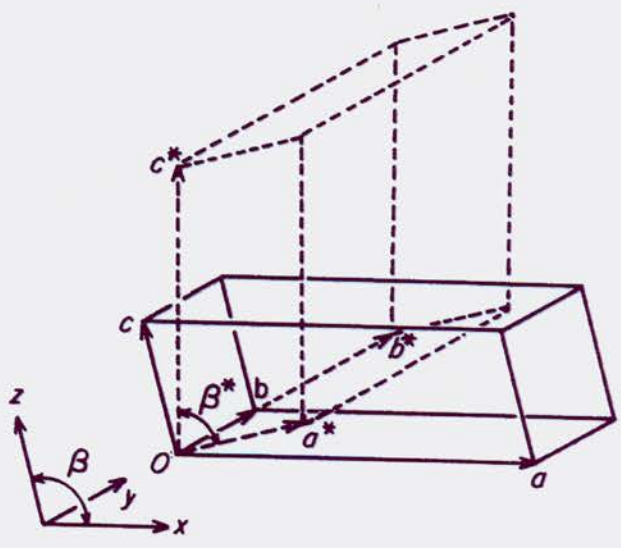
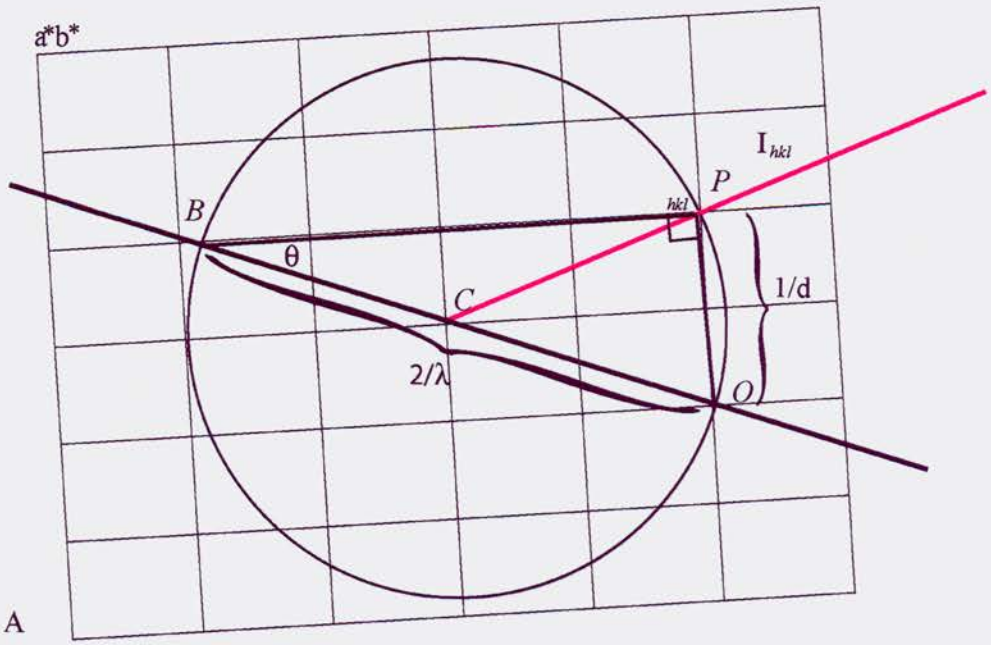
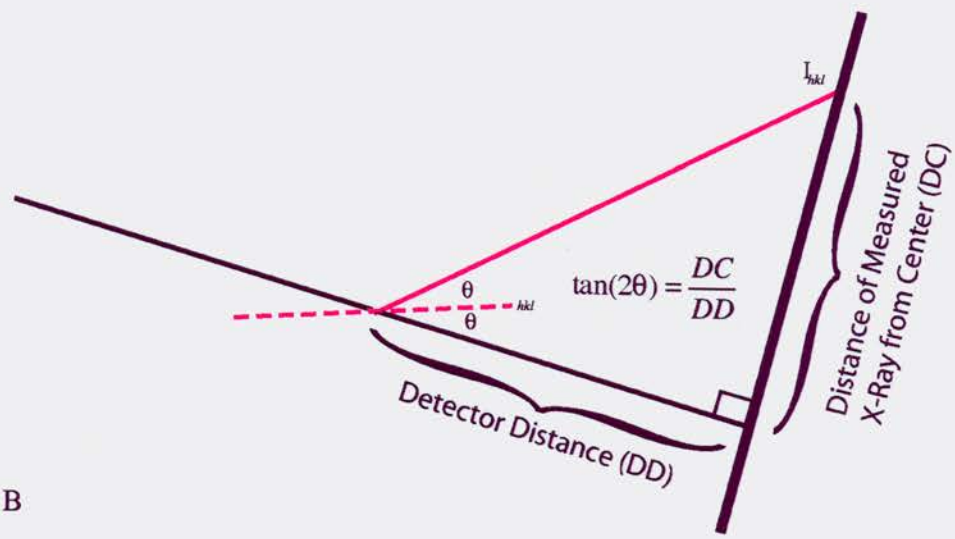

c*
c
b*
β*
b
z
β
y
x
O
a*
a

Figure 10. Prediction of an X-ray path. A. A two-dimensional diagram illustrating the mathematical description of the Ewald sphere. An x ray bisects a circle with B, C, and O lying on the circumference, the center, and the opposite circumference of the circle. The diameter of the circle is $2/\lambda$. O is an arbitrary origin of the reciprocal space lattice. P is a reciprocal lattice grid point that intersects with the semicircle. The pink line is the resultant X-ray. B. A diagram illustrating how the distance of a measured intensity can be related to θ.


a*b*
I_{hkl}
P
hkl
B
θ
C
1/d
2/λ
O

A


I_{hkl}
θ
θ
hkl
$\tan(2\theta) = \frac{DC}{DD}$
Detector Distance (DD)
Distance of Measured X-Ray from Center (DC)

B

Figure 11. Ewald's sphere. Three dimensional diffraction from the crystal.

Reproduced from Stout, G. H., and Jensen, L. H. (1989) X-ray Structure Determination: A Practical Guide, 2nd Edition.

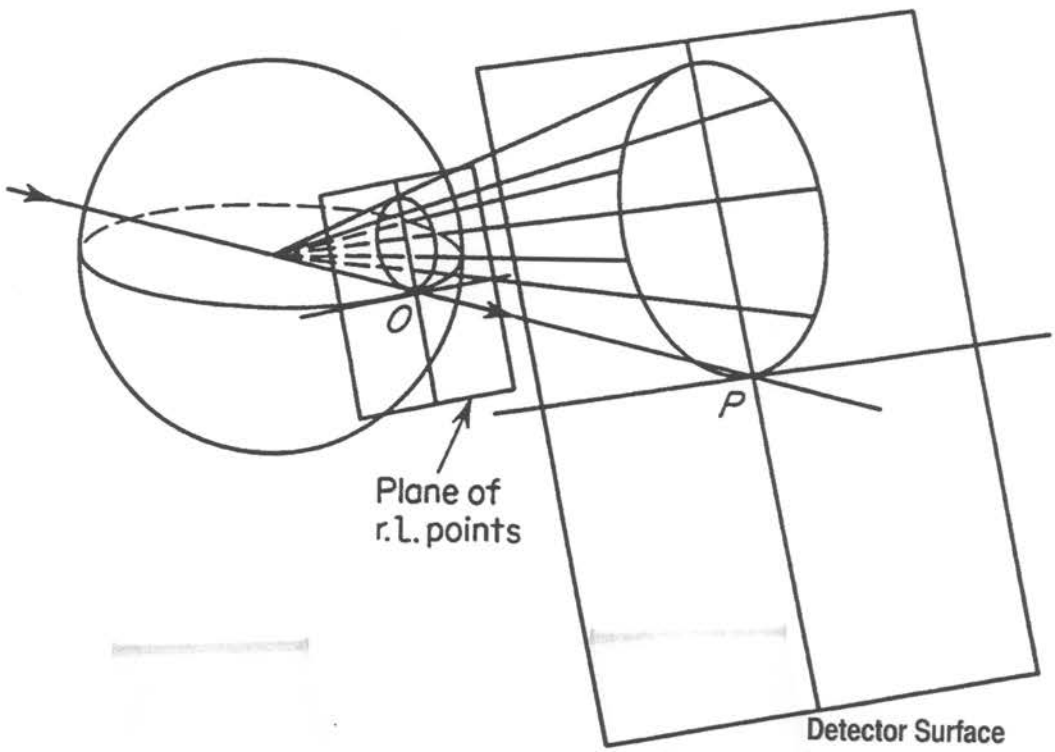

O
P
Plane of
r.l. points
Detector Surface

Figure 12. Crystallization phase diagram. The crystal growth requires the solution to be supersaturated as labeled at the top of the phase diagram. However, crystal nucleation only occurs under the labile conditions of the phase diagram. To avoid overnucleation the conditions must be such that they in the labile portion of the phase diagram close to the metastabile portion. At this position in the phase diagram, nucleation of a crystal will lower the protein concentration into the metastable region where crystal growth will occur but crystal nucleation will not.

Reproduced from the graduate dissertation of Joy Huffman from the Department of Biochemistry at Oregon Health & Sciences University. Her diagram was modified from McPherson, A. (1999) Crystallization of Biological Macromolecules.

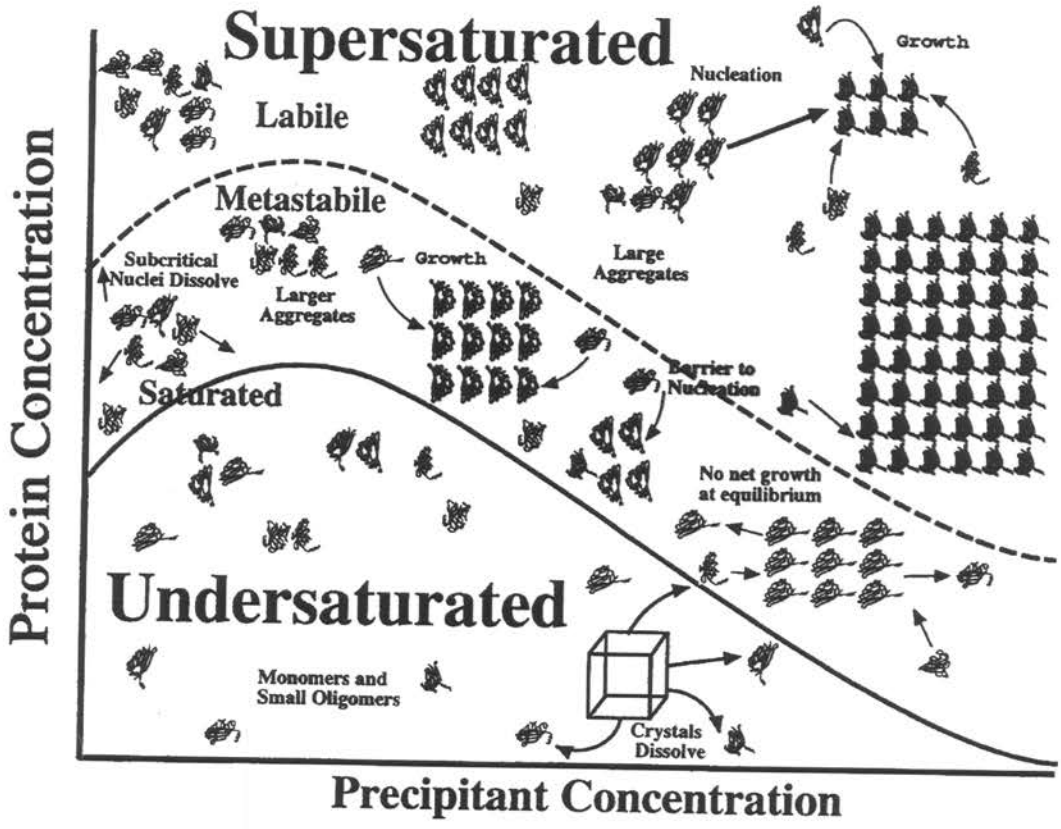
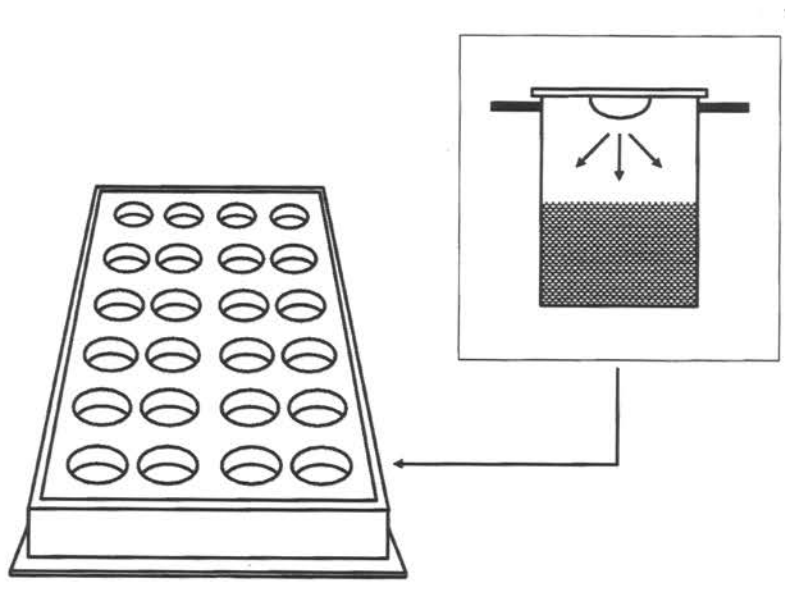
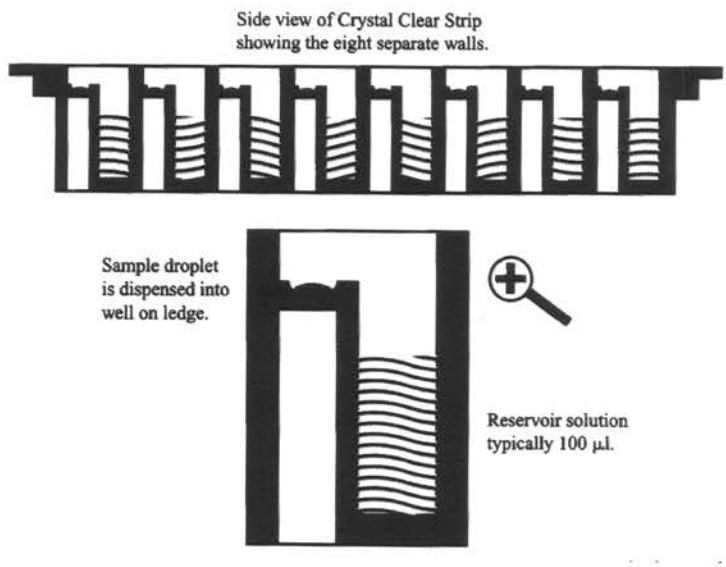

Supersaturated
Labile
Metastabile
Saturated
Undersaturated
Nucleation
Growth
Large Aggregates
Subcritical Nuclei Dissolve
Growth
Larger Aggregates
Barrier to Nucleation
No net growth at equilibrium
Monomers and Small Oligomers
Crystals Dissolve
Protein Concentration
Precipitant Concentration

Figure 13. Examples of crystallization trays. A. Hanging drop crystallization tray. B. Sitting drop crystallization tray.

Reproduced from McPherson, A. (1999) Crystallization of Biological Macromolecules.



A


Side view of Crystal Clear Strip
showing the eight separate walls.
Sample droplet
is dispensed into
well on ledge.
Reservoir solution
typically 100 µl.

B

Figure 14. Transition metal emission and absorption spectra. A. Copper emission spectrum (solid) and copper absorption spectrum dotted. B. Copper emission spetrum (solid) and nickel absorption spectrum.

Reproduced from Rhodes, G. (2000) Crystallography Made Crystal Clear, Second Edition.

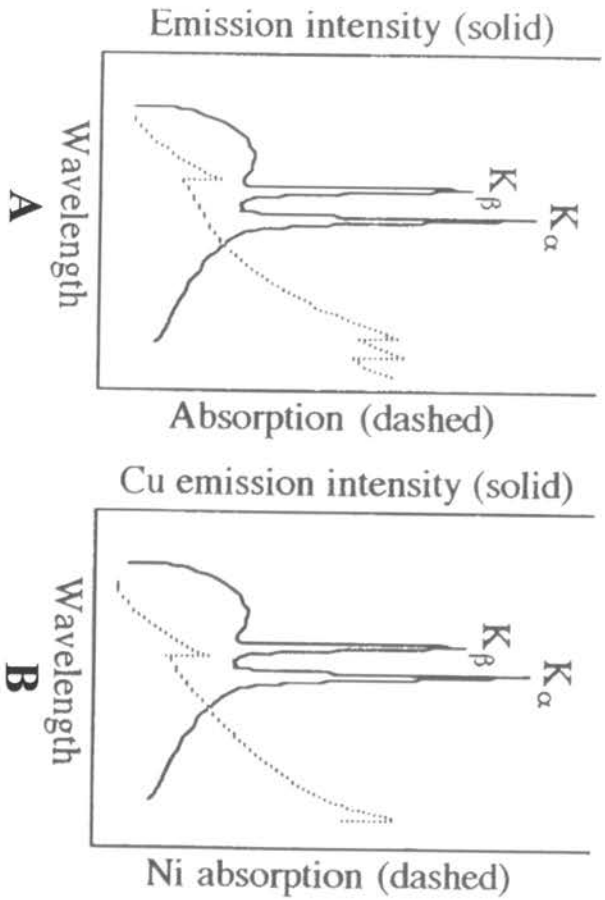

Emission intensity (solid)
Wavelength
A
K_β
K_α
Absorption (dashed)
Cu emission intensity (solid)
Wavelength
B
K_β
K_α
Ni absorption (dashed)

Figure 15. Copper emission spectrum filtered with nickel

Reproduced from Stout, G. H., and Jensen, L. H. (1989) X-ray Structure Determination: A Practical Guide, 2nd Edition.

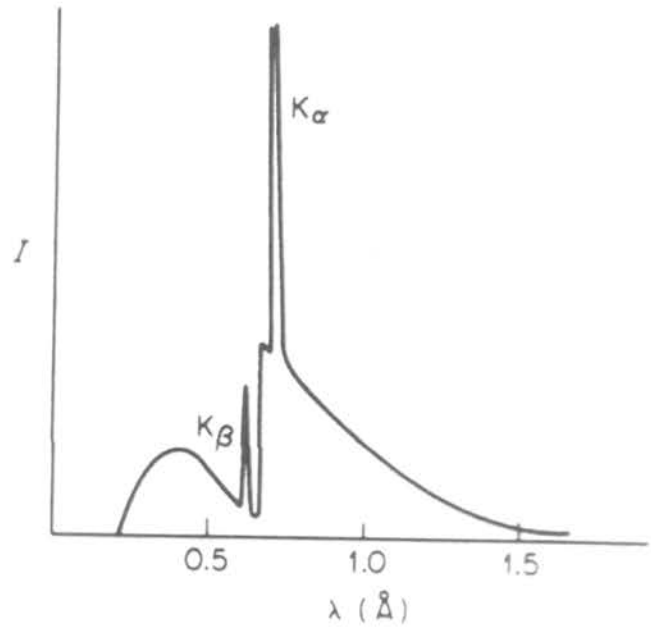

I
K_α
K_β
0.5
1.0
1.5
λ (Å)

Figure 16. Crystal twinning. Red and blue colored sections of the two crystal domains illustrates the difference in the orientation of the domains.

Adapted from Yeates, T. O. (1997) Detecting and overcoming crystal twinning, *Methods Enzymol 276*, 344-358.

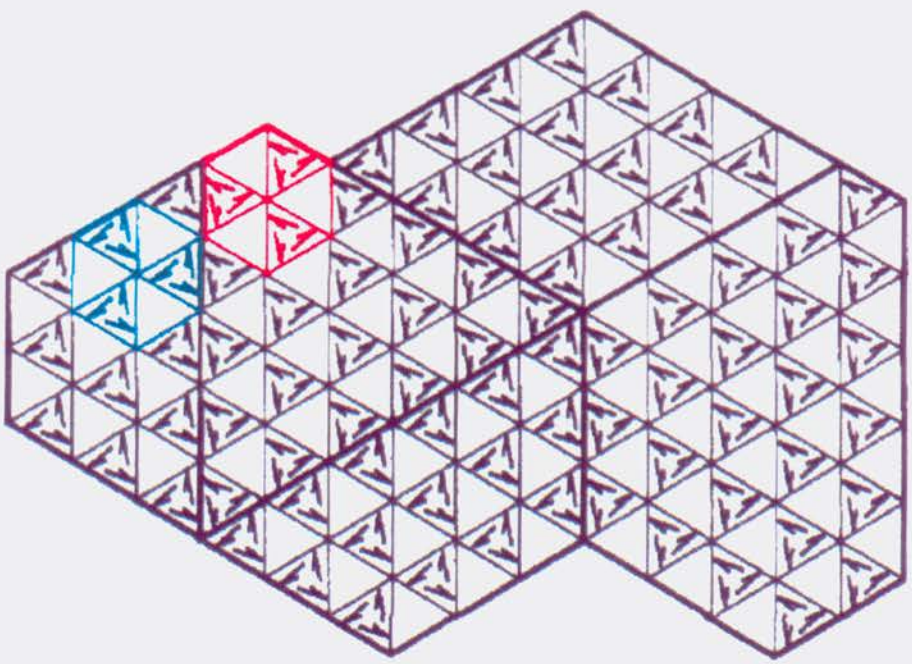
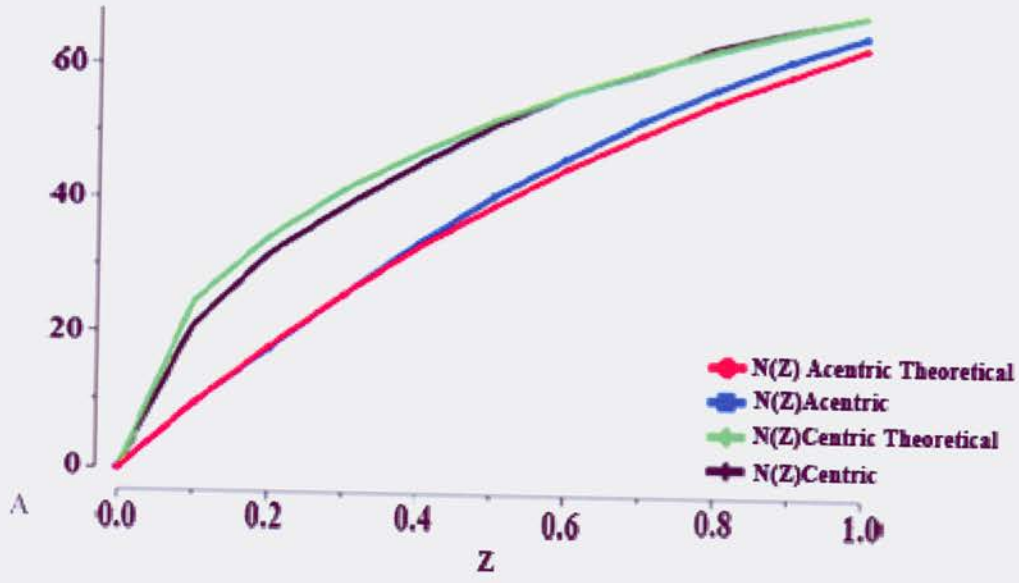


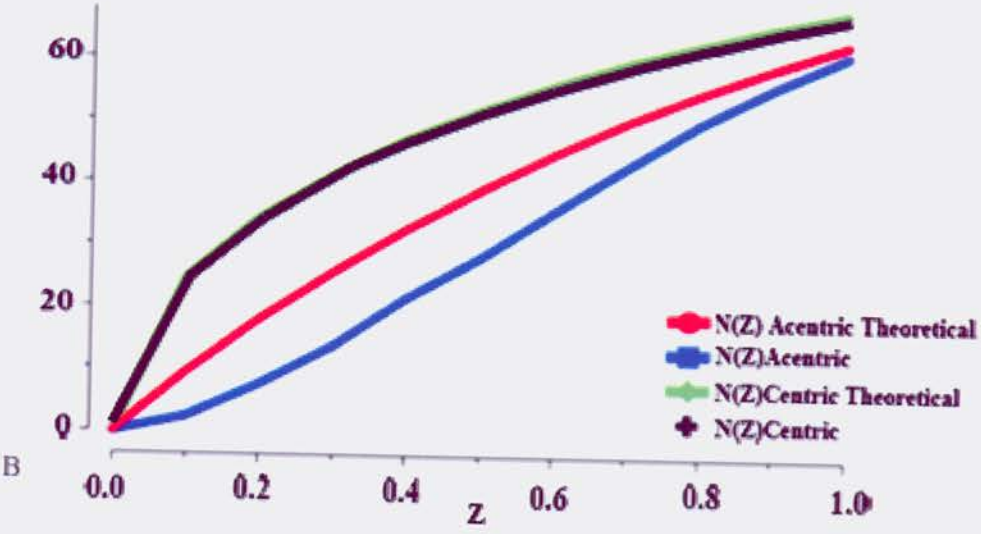
Figure 17. Detecting twinning with the cumulative intensity distribution. The cumulative intensity distribution is a plot relating the percent of the reflections that have a particular normalized intensity. A. The normalized plot of the distribution of the centric and acentric reflections is slightly bowed in each case. B. When twinning is present the plot of the acentric distribution is sigmoidal. This is due to the fact that the intensity is less random, as the sum of two different reflections due to the overlap of the reflections from two crystal domains are likely to be stronger. Thus the plot is weighted towards the higher normalized intensities when twinning is present.

Reproduced from the CCP4 web site:

http://www.ccp4.ac.uk/dist/html/pxmaths/bmg10.html#figurew

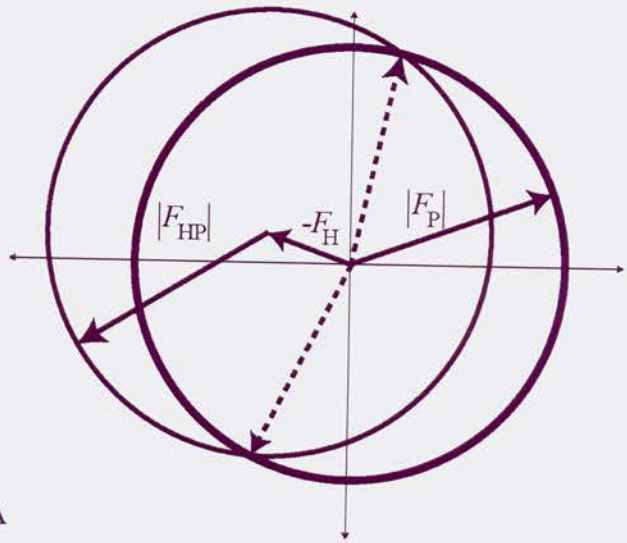

60
40
20
0
0.0
0.2
0.4
0.6
0.8
1.0
Z
N(Z) Acentric Theoretical
N(Z)Acentric
N(Z)Centric Theoretical
N(Z)Centric

A


60
40
20
0
0.0
0.2
0.4
0.6
0.8
1.0
Z
N(Z) Acentric Theoretical
N(Z)Acentric
N(Z)Centric Theoretical
N(Z)Centric

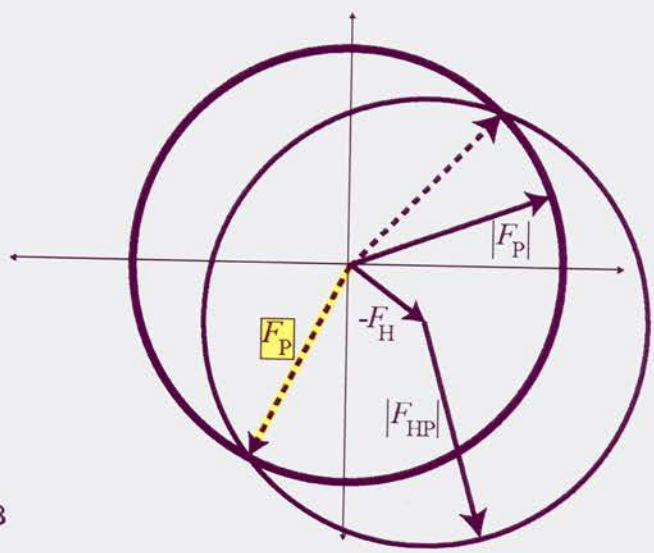
B

Figure 18. Argand diagram representation of MIR phasing. A. This diagram represents the phase possibilities determined by the first protein-heavy-metal derivative in MIR phasing. B. This diagram represents the final phase solution determined by the second protein-heavy-metal derivative.



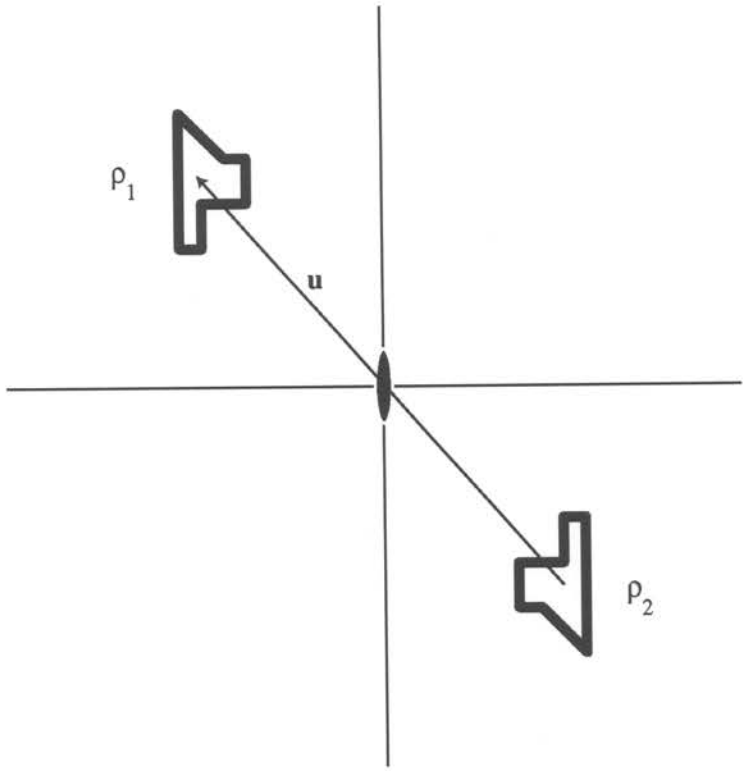
A

= Possible F_P Solution



B

Figure 19. Translation function. The intermolecular Patterson vector **u** relates the symmetrically related electron density from the model (ρ_1) to the symmetry related copy (ρ_2). In this case the symmetry is a two-fold axis as represented by the black oval.


ρ_1
u
ρ_2

Chapter 3: Methods

Mutagenesis and Expression Constructs

Wild type and genetically modified QacR genes in the recombinant expression vector pTTQ18 were used to produce protein for these experiments (*69*). For better recombinant expression in *E. coli*, all QacR expression constructs were codon-optimized and two non-essential cysteines were modified (C72A/C141S), with no effect on activity (*40*). Furthermore, a C-terminal hexahistadine tag was added for purification of the protein by the use of nickel-nitrolotriacetic acid-agarose affinity resin (Ni-NTA) (Qiagen, Valencia, CA). Glutamates 90 and 120 were individually substituted by PCR in Ronald Skurray's laboratory. Automated DNA sequencing, performed at the Australian Genome Research Facility verified the correct sequences. Furthermore, the substitution of lysine 67 to serine, which was shown to aid in crystallization, was added to both of the 90- and 120-glutamine-substituted QacR genes through the quick-change mutagenesis protocol. The K67S substitution had no observable change in affinity for R6G or DNA as measured by fluorescence polarization (data not shown). The following primers were used in the lysine 67 to serine mutagenesis:

forward- GGCAAGAACAGTGGAAATCGGAACAAATCAAAGC;

reverse- GCTTTGATTTGTTCCGATTTCCACTGTTCTTGCC.

The constructs were verified by sequencing at the Vollum Sequencing Core Facility

Protein Expression

Protein expression was induced in DH5α *E. coli* with 0.5 mM IPTG in the mid-log growth phase (0.4 – 0.8 O.D.). After 3-5 hrs of incubation, cells were harvested and

stored at -20°C. Ten milliliters of thawed cells were brought up in buffer A (50 mM Tris, 50 mM imidazole and 300 mM NaCl with a final pH 7.5) to 30 mL. After the addition of 500μg of DNAse I, the cells were lysed by French press at 4°C. The supernatant was then placed over a Ni-NTA column pre-equilibrated with buffer A. Next, the column was washed first with 100 mL of buffer A and then with 100 mL of buffer A with the imidazole concentration adjusted to 100 mM. The protein was eluted in 5 mL fractions with buffer A with the imidazole concentration adjusted to 1 M. The purity was assessed by SDS-polyacrylamide gel electrophoresis and coomassie staining, and the best fractions were pooled.

Crystallization

For reproducible crystallization of the induced conformation of QacR, reductive alkylation of the lysines of the purified protein was necessary (*1; 70*). The extra imidazole from protein purification and the by-products from reductive alkylation were removed by buffer exchanged 3 times with 1:10 dilutions into buffer A. The purified QacR was incubated with 50 - 500 μM ligand and clarified by centrifugation before setting-up hanging drop crystallization experiments. The ligands included R6G, ethidium bromide, MG, Dq, DB 75, and DB 359. The 10 μL drops included equal volumes of protein with ligand in wash buffer and reservoir buffer (2.7 - 2.3 M ammonium sulfate and 0.1 M sodium acetate pH 4.6 – 5.0). Crystals took between 4 days and two weeks to reach full size and grew to a maximum of 400 μm in all directions in one of two different space groups: $P4_22_12$ or $P6_2$. The $P4_22_12$ space group crystals usually grew as rectangular prisms with the dimensions 100 x 100 x 200 but sometimes grew to close to twice the size. The $P6_2$ space group crystals grew as hexagonal rods with similar

dimensions as the tetragonal rods. The $P6_2$ crystals grew most often as two diamond-shaped sections sharing one surface with a maximum size of 100 – 200 μM.

Data Collection

All X-ray diffraction data were collected at 100 K at the Advanced Light Source (ALS) in Berkeley, CA and the Stanford Synchrotron Research Laboratory (SSRL) in Palo Alto, CA. To protect the crystals from freezing at these temperatures, they were soaked in a cryoprotectant containing 1.6 M ammonium sulfate, 0.1 M sodium acetate, and 20% glycerol. The X-ray diffraction data from the E120Q(K67S)-QacR-R6G, E120Q-QacR-Dq, E90Q(K67S)-QacR-R6G, and E90Q(K67S)-QacR-Et complex crystals were collected on the beam line ALS 4.2.2 (Molecular Biology Consortium) using the Blu-Ice control program collect data and d*trek (MSC Inc., Woodlands, TX) for integration and scaling (*71; 72*). The X-ray diffraction data from the E90Q-QacR-MG and E90Q(K67S)-QacR-Et complex crystals were collected at ALS 8.2.2 (Howard Hughes Medical Institute) and the E90Q-QacR-Dq data were collected as ALS 8.2.1 (Howard Hughes Medical Institute). These data sets were collected using the Blu-Ice control program and data integration and scaling were performed with MOSFLM and SCALA from the software package CCP4 (*68*).

Phasing and Refinement

Molecular replacement in CCP4 using QacR-dequalinium as a search model was used to solve the structures of the ligand-QacR complexes (*68*). The ligand coordinates as well as the related stereochemistry files for use with the program O and topology and parameter files used with CNS (Crystallography NMR System), were obtained from the

Hetero-compound Information Centre – Uppsala (Hic-Up server; http://alpha2.bmc.uu.se/hicup/) (*67; 73; 74*). The program O was used to visualize and manipulate the structures for model building and CNS was primarily used for refinement of the crystal structures. To check for crystal twinning and estimate the degree of twinning in the $P6_2$ space group crystals, structure factor amplitudes were analysed with the Merohedral Crystal Twinning Server created by Todd Yeates (http://nihserver.mbi.ucla.edu/Twinning/) (*57*). Scripts in CNS were used to deconvolute the twinned data, based on the twinning estimate from the twinning server. The detwinned data were treated normally thereafter.

All model refinement consisted of an initial round of rigid body refinement followed by two rounds of simulated annealing refinement. Positional refinement followed all further adjustments to the structures. The subsequent adjustments to the structure, including the addition of sulfate and water molecules, were made by rebuilding into the $2F_o$-F_c and F_o-F_c electron density maps using the software package O. In addition to van der Waals and hydrogen bonding distance constraints, the following overlapping electron density criteria were used for solvent, ion, and ligand placement: $2F_o$-F_c > 1.0 σ and F_o-F_c > 3.0 σ. Simulated annealing composite omit maps ($2F_o$- F_c) were used to verify the final model. Models were validated with the "model check" script from CNS and with PROCHECK in the CCP4 suite of programs.

Data analysis was performed through comparison of three-dimensional structural alignments of QacR crystal structures. Alignments of structures were performed in O using the least-squared fitting function (LSQ-EXP/MOL). Interatomic distances were determined using the distance measurement tool in O. The center-to-center distances of

the recognition helices (α3) were measured by averaging the positions of the αC of the residues of a turn in the middle of each recognition helix: residues 37, 38, 39, and 40. The distance between each of these points was determined by the formula

$$\text{distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \ .$$

To compare twisting motions of the DNA binding domains, the relative orientations of the recognition helices was compared to their orientation in the WT-QacR-DNA complex structure. An alignment of the B subunits of apo dimer from the WT-QacR-Dq structure, the Dq bound dimer from the same structure, and the MG bound dimer from the E90Q-QacR-MG complex structure with the C subunit of the DNA bound structure was performed. Vectors spanning from one recognition helix to the other were measured for each structure and measured with the following formula:

$$\text{Angle between vectors} = \cos^{-1} \frac{x_1 \cdot x_2 + y_1 \cdot y_2 + z_1 \cdot z_2}{\sqrt{x_1^2 + y_1^2 + z_1^2} \cdot \sqrt{x_2^2 + y_2^2 + z_2^2}}$$

Volume changes of the QacR binding pocket were determined with VOIDOO (*75; 76*). The volume was measured as the solvent accessible volume of the cavity using a 1.4 Å probe.

Affinity Measurements

Fluorescence Quenching

The majority of the ligand binding data included in this thesis were measured with fluorescence-quenching experiments performed by Katie M. Hardie in Ronald Skurray's laboratory in Sydney, Australia. These experiments were performed previously in order

to determine QacR ligand affinities and QacR induction characteristics for various dyes, antiseptics, and plant alkaloids (*6*).

Fluorescence Polarization

Fluorescence of polarization was used to test for R6G affinity. Polarization measures the difference in the rate of rotation of a population of fluorescent molecules in a solution. As free R6G rotates very quickly compared to protein-bound R6G, there is a large difference in the polarization signal between free and bound R6G that may be used for measuring R6G by the protein. Polarization is measured by comparison of the intensities of fluorescently emitted polarized light that is perpendicular and parallel to the polarized plane of the excited light.

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

where $I_{\parallel}$ is the intensity of the parallel component of the emitted light and $I_{\perp}$ is the intensity of the perpendicular component of the emitted light (*77*).

The Beacon 2000 (Invitrogen) fluorescence polarization machine was used to measure polarization. Excitation and emission wavelengths were selected with two 10 nm band-pass filters selecting for 530 nm and 560 nm light, respectively. Experiments were performed at 25 °C, and a concentration of 3 nM R6G was used in 1 mL of either the buffer A or the buffer B (100 mM NaCl, 25 mM Tris pH 7.5, 2.5% glycerol). Protein was added from a 100 μM stock. The resulting data were analyzed using Kaleidagraph (Synergy Software).

The dramatic reduction in fluorescence intensity observed over the course of the experiment affects the polarization data and was corrected with the following formula.

$$R = \frac{\{[(P - P_f)/(P_b - P)](Q_f/Q_b)(P_b)\} + P_f}{1 + [(P - P_f)/(P_b - P)](Q_f/Q_b)}$$

where R = corrected polarization values; P = measured polarization; P_f = polarization, when ligand is totally free; P_b = Polarization, when ligand is totally bound; Q_f = intensity, when ligand is totally free; and Q_b = intensity, when ligand is totally bound. After plotting the protein concentration against the polarization, the K_d is determined by fitting in Kaleidagraph with the following hyperbolic equation:

$$P = \left(\frac{(P_b - P_f)[\text{protein}]}{K_d + [\text{protein}]}\right) + P_f$$

Isothermal Titration Calorimetry (ITC)

Dequalinium, DB75 and DB359 affinity measurements were made with ITC. All protein samples were extensively dialyzed in buffer A and dequalinium or other ligands were dissolved in the dialysis buffer. Ligand and protein concentrations were determined spectrophotometrically (Table 1). Protein and ligand samples were degassed before they were loaded into the cell and syringe of the VP-ITC from MicroCal Inc. (North Hampton, MA). The stirring speed was set to 300 r.p.m. and the power was set to 10 μcal/second. The experiment was set up with the protein solution in the cell and the ligand in the syringe and in reverse. Data analysis was performed with the Origin 5.0 package from MicroCal Inc. These experiments directly measure ΔH and K_a from which ΔG and ΔS can be calculated. The stoichiometry of the interaction, n, is also determined by ITC. Further reading concerning ITC can be found with the following references (*78-80*).

Fluorescence Polarization Competition Assay

The dequalinium affinity was confirmed by competition with R6G for the protein. The experiment was set up similarly to the previous fluorescence polarization experiments, with 3 nM R6G but in this case, the R6G was pre-bound to the protein (~3 μM) and competed off with successive additions of the other ligand. The polarization data are corrected for the intensity change in the experiment as described in the previous fluorescence polarization method. After plotting the protein concentration against the polarization, the EC_{50} is determined by fitting the following sigmoidal equation:

$$P = P_f + \frac{(P_b - P_f)}{(1 + e^{C - EC_{50}/H})}$$

where C is the concentration of the competitive drug; EC_{50} is the midpoint of the competition; and H is the hill slope. The EC_{50} is easily converted to the K_d of the ligand with the following equation: $K_d = EC_{50}/(1+([ligand]/K_d)$.

Table 1. Table of extinction coefficients

Drug	Solvent	Ext. Coeff.	Abs./Ex.	Em.	Reference
Dq		23,400	329	360	
EtBr		5,500	482	616	MP & Σ
R6G	EtOH	115,948	529.5	552	(*81*)*
MG		148,900	616.5	nil	(*81*)*& MP
RhB	EtOH	106,000	542.8	565	(*81*)*
R123	EtOH	85,200	511.8	535	(*81*)*
	MeOH	101,000	507	529	MP
R110	pH7	83,000	496	520	MP
	MeOH	92,000	499	521	MP
Fluorescein	pH9	93,300	490	514	MP
DB75		36,600	360	462**	DB
DB359	0.1M NaCl MES buffer	28,570	327	456**	DB

* – http://omlc.ogi.edu/spectra/

** – 100mM NaCl, 15mM NaCl, and 2.5% Glycerol pH 7.5

MP – Molecular Probes/Invitrogen

DB – David Boykin's Laboratory (Fluorescence spectra determined by author)

Σ – Sigma-Aldrich

Chapter 4: X-ray Crystallography Reveals New Binding Modes for Ethidium Bromide and Rhodamine-6G with a Glutamate to Glutamine Substituted QacR at Position 90

Introduction

Glutamate 90, as one of the four glutamates in the larger QacR ligand-binding pocket, appears to be critical for binding R6G, ethidium, and MG. Crystal structures of wild-type QacR ligand, WT-QacR-ligand, complexes indicate that E90 is proximal of all of these ligands (*1*). Like all QacR ligands, these molecules are cationic, thus the anionic nature of E90 may be essential for interactions with these ligands. To determine the degree to which the formal negative charge of the residue aids in ligand binding, we substituted glutamine at this position and studied the interactions of this substituted protein with the structurally characterized QacR ligands: ethidium, MG, Dq, and R6G. The affinities of the substituted QacR to each ligand were measured with tryptophan fluorescence quenching, fluorescence polarization, and isothermal titration calorimetry. Interactions of these ligands with the E90Q substituted QacR (E90Q-QacR) were also studied structurally with X-ray crystallography.

The data indicate that the formal negative charge of glutamate 90 is not critical for ligand affinity. None of the affinities were adversely affected by glutamine substitutions at this position. However, two of the ligands are seen in alternate positions in the crystal structures suggesting that there are multiple sites in the binding pocket with similar affinities.

Results

Ligand Affinity

Substitution of E90 with glutamine, a residue without a formal negative charge, does not greatly affect ligand binding (Table 1). The affinity for DNA was that of wild type, which illustrated that the mutant protein was intact and functional. The binding affinities of E90Q-QacR for MG and R6G remain unchanged from the unsubstituted QacR. The binding affinity of E90Q-QacR for ethidium is two-fold greater than that of WT-QacR (1.18 ± 0.04 μM to 2.4 ± 0.1 μM). The binding affinity of E90Q-QacR for Dq is two-fold less that of WT-QacR (2.2 ± 0.6 μM to 1.0 ± 0.5 μM). However, the difference in the Dq affinity is close to the error of the measurements so this does not appear to be significant (Figure 1). The stoichiometry varied and it was not clear why it was not equal to 1.

Crystallization

The ease and characteristics of crystallization of the E90Q-QacR protein depended significantly on the ligand with which it was cocrystallized. E90Q-QacR-MG was the easiest complex to crystallize and diffracted to the highest resolution, 2.2 Å, which is better than any of the wild-type crystals as well. Interestingly, the E90Q-QacR-MG complexes crystallized in $P6_2$ space group rather than $P4_22_12$ space group crystals observed with the WT-QacR-ligand complexes (Table 2) These crystals formed pairs of diamond shaped hexagonal prisms, fused back-to-back (Figure 2A). The $P6_2$ E90Q-QacR-MG complex crystals have only one dimer in the asymmetric unit. In comparison, the asymmetric unit of the $P4_22_12$ WT-QacR-MG complex crystals has two dimers: a

drug bound dimer and an empty or apo dimer (WT-QacR-Apo). The E90Q-QacR-Dq complexes also crystallized in the $P6_2$ space group but could also form crystals in the $P4_22_12$ space group observed by the WT protein. In this case, the $P6_2$ crystals diffracted to a much higher resolution then the $P4_22_12$ crystals, but the electron density for Dq was not satisfactory and α6 in the binding pocket appeared less ordered. Thus, the data from the $P6_2$ crystals were not used in the analysis of the binding pocket.

The E90Q-QacR-R6G and the E90Q-QacR-Et crystals were much more difficult to obtain and these complexes did not crystallize without a K67S substitution. As the $P6_2$ space group crystals of E90Q-QacR-MG and E90Q-QacR-Dq diffracted to the highest resolution observed for QacR, it was hypothesized that inducing $P6_2$ space group crystals would increase the resolution of the subsequent QacR structures. Analysis of the E90Q-QacR-MG and WT-QacR-Dq complexes indicated that there was a crystal contact involving K67 that was unique to the $P4_22_12$. It was hypothesized that substituting the surface amino acid K67 with a serine might induce crystallization in the higher resolution $P6_2$ space group observed with E90Q-QacR-MG complex crystals. Although, E90Q(K67S)-QacR-ligand crystals diffracted no better than the WT crystals observed previously, this substitution allowed for the crystallization of the E90Q-QacR-R6G and the E90Q-QacR-Et complexes. The E90Q(K67S)-QacR-Et complex formed exclusively $P6_2$ crystals as expected. Surprisingly, however, the E90Q(K67S)-QacR-R6G complexes formed both $P6_2$ and $P4_22_12$ crystals, and the best data were collected from the latter (Figure 2B&C). However, the tetragonal crystals could be epitaxially twinned and these crystals could not be used for data collection (see Crystal Twinning section in Chapter 2)(Figure 2C). Furthermore, the $P6_2$ K67S crystals did not always yield higher

resolution. The E90Q(K67S)-QacR-Dq formed large hexagonal crystals easily, but these crystals did not diffract well and were epitaxially twinned (Figure 2D). Twinning could sometimes be overcome, as the E90Q(K67S)-QacR-Et complex crystals were also epitaxially twinned. After testing a number of crystals that could not be processed, one E90Q(K67S)-QacR-Et complex crystal produced a data set that could be processed correctly despite the twinning (Figure 2E). As the K67S substitution has no effect on the activity of the protein the substitution is not included in the nomenclature of the substituted proteins. However, the table of crystallographic and model statistics for the various complex crystals includes whether the substitution was needed in each particular case.

Please refer to Table 2 for further crystallographic and model statistics.

E90Q-QacR-MG complex structure

The 2.2 Å resolution structure of the E90Q-QacR-MG complex reveals ligand binding is the same as the WT-QacR-MG complex structure, however the conformations of these structures are different. The portions of the protein that interact with MG overlay with a 0.7 Å RMSD, and the MGs from the structures are in identical positions in the binding pocket in each structure (Figure 3). However, the crystal structure of the E90Q-QacR-MG complex appears to be in a more condensed conformation than the WT-QacR-Dq complex structure (Figure 4). The two DNA binding domains are closer to each other and in consequence the E90Q-QacR-MG complex structure buries an extra 300 $Å^2$ compared to the WT-QacR-Dq complex structure, which is the highest resolution structure of drug bound QacR in the $P4_22_12$ space group.

The distance between the recognition helices is smaller in the $P6_2$ E90Q-QacR-MG complex structure than the distance of the recognition helices in the $P4_22_12$ structures. The recognition helices are 39.1 Å apart in the E90Q-QacR-MG complex structure, compared to 46.4 Å in the $P4_22_12$ WT-QacR-Dq structure (Figure 5). The distance of the helices on the E90Q-QacR-MG complex structure is the same as that of the recognition helices from the WT-QacR-DNA complex structure (39.2 Å). Furthermore, the distance of the helices on the WT-QacR-Apo structure (the apo dimer in the asymmetric unit of the WT-QacR-Dq complex structure) is also the same, at a distance of 39.4 Å.

However, the twisting motion of the A subunit on ligand binding is similar between the $P6_2$ and the $P4_22_12$ structures. The alignment of the B subunits of the WT-QacR-Apo structure, WT-QacR-Dq complex structure, and the E90Q-QacR-MG complex structure with the C subunit of the DNA bound structure revealed angular changes in the orientations of the recognition helices that occur on ligand binding (Figure 6A&B). Compared to the orientation of the recognition helices in the WT-QacR-DNA complex structure, the recognition helices of the WT-QacR-Dq complex structure are set at an angle of 26° (Figure 6C). In comparison the recognition helices from the E90Q-QacR-MG structure are only set at an angle of 21°.

The E90Q-QacR-MG complex structure also revealed a few molecules in the binding pocket not previously identified in the WT-QacR-MG complex structure. Otherwise the structure of the ligand-binding pocket is the same as that of the WT-QacR-MG complex. The new electron density in the pocket is best fit with imidazoles. (Figure 7, Table 3).

E90Q-QacR-Dq complex structure

The position of the Dq in the 3.3 Å resolution structure of the E90Q-QacR-Dq complex is similar to the Dq in the WT-QacR-Dq complex structure. In the portion of the structure proximal to E120, the Dq is positioned exactly the same way. However, closer to E90, the aminomethylquinolinium head group of the Dq appears to twist to a position 25° from the wild-type position (Figure 8). The head group also curls in slightly toward the other head group. Relative to E90, the carboxamide of the Q90 side chain moves 3.6 Å to a position closer to the head group of the Dq. In the wild type, the carboxylate of E90 is positioned 3.7 Å away from the ε-nitrogen of histidine 164. These residues appear to be joined by a salt bridge or hydrogen bond in the WT-QacR-Dq complex structure. For further comparison of contacts and distances between Dq and E90Q- and WT-QacR refer to Table 4.

Glutamine 90 appears to interact with both the ligand and Q96. The closest approach of the ε-nitrogen towards the aminomethylquinolinium ring system is about 3.2 Å. The ε-oxygen of E90Q comes within 3.1 Å of the ε-oxygen of Q96. The carbonyls are nearly parallel. The interaction of E90Q with the ligand is weak. The ligand has moved away from W61 slightly, as well. However, these changes in interactions are small, and if these changes in the protein ligand interactions are significant, they are opposing. It is not surprising that there is no net change in ligand affinity.

It is possible that after the loss of the salt bridge with H164, E90Q swings into another position that has more favorable interactions. In the WT-QacR-MG and WT-QacR-Et structures, the glutamate appears to be in a similar position to that of the glutamine. In these two cases, the glutamate interacts with the ligands in the structure.

Thus, the glutamate may be subject to competition between interacting with ligand and interacting with the histidine as a salt bridge. However, the E90 from WT-QacR-R6G shows that interacting with both ligand and histidine is possible. The position of R6G is closer to E90 in its position near the histidine and thus E90 does not require reorientation in order to interact with H164. Thus it appears that E90 prefers to form a salt bridge with H164. In the absence of this interaction or with the appearance of a positively charged ligand in the pocket near enough to interact with E90, the glutamate will reorient towards the ligand.

E90Q-QacR-Et complex structure

The ethidium in the 2.8 Å resolution structure of the E90Q-QacR-Et complex moves significantly compared to the position of the ligand in the WT-QacR-Et complex structure. The new position for ethidium appears to be at an angle of approximately 120° compared to the orientation of the ethidium in the wild-type structure (Figure 9). Furthermore, the center of mass moves away from Y103 towards W61 by 5 Å. This decreases the distance of ethidium to W61 from 6.5 to 3.9 Å for a cation-pi interaction between N23 and the tryptophan.

A number of ligand-residue contacts are gained and lost with the repositioning of ethidium (Table 5). The hydroxyl oxygen of S86 contacts the phenanthridinium ring system of ethidium along the edge where it is more positively charged (Figure 10). On the other side of the S86 hydroxyl group, a 2.9 Å hydrogen bond with the carbonyl of Y82 orients the serine residue for a polar interaction with ethidium. In another new dipole-charge interaction, the ∂-oxygen of N154 is positioned 3.3 Å from one of the two nitrogens that lie on either end of the phenanthridinium ring system and are the centers of

the positive charge in the molecule. A few residues with new van der Waals interactions with ethidium include M116, L119, and N154. The lost contacts include van der Waals interactions with the I100 and I99 as well as a contact between the carbonyl oxygen from F162' and one of the two nitrogens that lie on either end of the phenanthridinium ring system.

A number of ligand-residue contacts are similar between the repositioning of ethidium. Tyrosine 103 moves to a new position of the ethidium to maintain contact with the ligand. In the WT-QacR-Et complex structure, the ring systems of the ethidium and Y103 are face to face at a distance of 3.3 Å (Figure 9). After rotating around the α-β bond by about 60°, the tyrosine maintains a distance of 3.6 Å and has adopted a perpendicular orientation to the phenyl ring of the ethidium which is more typical of a pi-pi interaction. Glutamate 120 also repositions to accommodate the new position of ethidium and maintain a charge-charge interaction with the ligand. However, the distance of the glutamate to the ethidium moves from 4.0 Å from positively charged phenanthridinium ring system in the WT-QacR-Et complex to 3.3 Å from the ring system in the ligand position in the E90Q-QacR-Et complex (Figure 9). If the E120 remained in the original position, it would clash with the new position of the ethidium, as the closest approach would be within 1.6 Å. There is also a 2.6 Å change of the position of E90Q compared to E90 caused by a rotation around the α-β bond (Figure 9). This also decreases the distance between the residue and the ligand in comparison of the structures. The distance between E90 and the ethidium in the WT-QacR complex is 3.7 Å. The E90Q is positioned to participate in a dipole-charge interaction with the ethidium. The ε-oxygen of E90Q is only 3.3 Å from the positively charged phenanthridinium ring system

of the ethidium and 3.5 Å from one of its nitrogens where the positive charge is the most pronounced. Similar to the above examples, Q96, Y123, N157, and F162' have interactions with the ethidium in both structures.

Another interesting aspect of E90Q-QacR-Et structure is that there are patches of electron density on either side of the ethidium. Bromide ions fit well into these sites at distances of 3.7 Å from N23 of the ethidium and 3.7 Å from the edge of the phenyl ring of the ethidium. Both of these parts are positively charged. However, N23 should be more so. Both sites fit the bromides well individually, however the model fit is better if only the site proximal to the phenyl group is filled. This suggests that the ethidium can bring its own counter ion into the pocket. This counter ion could make the net charge of the ethidium zero and eliminate its dependence on charge-neutralization by the protein. Thus, counter ions could eliminate the need of the negative charge of the glutamates for cationic ligand recognition.

E90Q-QacR-R6G complex structure

In the 2.9 Å resolution structure of the E90Q-QacR-R6G complex, density for R6G appears in two mutually exclusive positions in the binding pocket (Figure 11). Position 1 for the rhodamine is almost exactly in the same position and makes similar contacts as the rhodamine in the WT-QacR-R6G complex structure (Figure 12A, Table 6). Position 2 of the rhodamine is in the ethidium binding pocket and the electron density in this portion is a little stronger than the density found in position 1 (Figure 12B). Furthermore, all of the amino acids are in the same orientations between the WT-QacR-R6G and the E90Q-QacR-R6G structures; even E90Q is in the same position as E90. Position 2 for R6G overlaps with position 1, thus making occupancy in the sites mutually

exclusive. Fluorescence polarization measurements determined that the stoichiometry of the ratio of E90Q-QacR dimer to R6G in the complex was approximately 1:1 (Figure 13).

Electron density for an additional orientation for glutamate 120 also suggests that the two binding positions of R6G are mutually exclusive (Figure 12B). Although there is density in the area of the structure corresponding to E120 of WT-QacR-R6G complex structure, this orientation is not compatible with position 2 of R6G. However, the F_o-F_c electron density suggests that the glutamate is also in an orientation for E120 compatible with position 2 of rhodamine. In this orientation, E120 contacts the N1 nitrogen and is parallel to the plane of the xanthylium ring system of the R6G at a distance of 3.5 Å. The glutamate makes a complementary charge interaction with R6G N1. Therefore, it appears that E120 is in the orientation of E120 from the WT-QacR-R6G complex structure when R6G is bound in position 1. However, when R6G is in position 2, this orientation for E120 is excluded and E120 moves to a position that has favorable interactions with the R6G.

The second position for rhodamine has minor differences for the positions of ethidium observed in previous structures. Considering that ethidium and R6G have similar structures consisting of a three-ring conjugated system and phenyl ring that extends from the side, position 2 of R6G is flipped over by 180° compared to the ethidium in the WT-QacR-Et complex structure (Figure 14A). Again, this would require reorienting of E120, as the orientation of the glutamate in the WT-QacR-Et complex structure is the same as the orientation in the E90Q-QacR-Et complex structure. The binding pocket is also different, as Y103 and Y123 are in different orientations when ethidium is bound to the wild-type QacR.

In the E90Q-QacR-Et structure, the ethidium is oriented closer to position 2 of R6G in the E90Q-QacR-R6G complex than the ethidium in the WT-QacR-Et complex. However, the ethidium is tilted out of the plane of the R6G by about 45° (Figure 14B). Interestingly, Y103 and E120 orientations are similar between the E90Q-QacR-R6G complex and the ethidium in the E90Q-QacR-Et, yet the orientation of Y123 in the E90Q-QacR-Et complex is not like that of the Y123 of E90Q-QacR-R6G and remains similar to that of the WT-QacR-Et complex. It is clear that the orientation of Y123 in the E90Q-QacR-R6G structure would clash with the position of ethidium in the E90Q-QacR-Et structure. Furthermore, of the two orientations of Y123, the orientation for Y123 seen in the WT-QacR-Et complex appears to have slightly better contacts for the phenyl ring of ethidium.

The contacts by the R6G in position 2, other than those with E120, are primarily van der Waals contacts. These van der Waals contacts are provided by S86, I99, V54, M116, I124, Q154, N157, and F162'. However, Y123 and Y103 appear to be making π-π interactions with the xanthylium ring system of R6G. Typical of π-π interactions the rings of these tyrosines are both approximately perpendicular to the xanthylium ring system of R6G and are 3.4 and 3.8 Å away, respectively.

Discussion

Role of E90Q for Ligand Affinity

These experiments indicate that a neutralization of a prominent binding site charge through an amino acid substitution in QacR does not adversely affect binding, yet

this substitution affects the binding positions of ethidium and R6G. The only change in affinity observed after substitution with an uncharged ligand was the two-fold increase of affinity for ethidium. Though this change is well above measured error, a two-fold difference is only a gain of 0.4 kcal/mol. Considering that a hydrogen bond is approximately 3 to 5 kcal/mol, this change might not be reflected structurally in an obvious way (*82*). However, π-cation and π-π interactions have energies in this range, so extra ligand-protein interactions with either of these types of interaction may account for the change (*83*). The changes in the ligand-residue interactions observed in the E90Q-QacR-ethidium structure involve aromatic residues and might account for this change in affinity. Regardless, a two-fold change is small.

The negative charge of E90 does not play a major role in the binding affinity of its cationic ligands. The tyrosines provide a flexible binding surface that can interact with ligands through π-π, cation-π, aryl-sulfur, hydrogen bonding, and van der Waals interactions (*83*). The π-electrons of these aromatic residues are partially negative and therefore may be all that is necessary to complement the positive charge of the ligand. There are also a number of hydrophobic and a few polar interactions available in the pocket. The interaction with E90 may be more important for molecules that do not take as much advantage of the other ligand binding interactions in the pocket.

The glutamate is also a part of the larger negative charge of the pocket, created not only by the other glutamates, but also by the main-chain carbonyls and the π-electrons of aromatic residues. The electronegativity of the pocket may be all that is required for binding of the ligand. It is difficult to bury an uncomplemented positive charge in the dielectric of a protein. The global negative charge might remove that

barrier to allow the interactions with the pocket to occur without specific interactions between negatively charged residues and the ligand. Further studies on the global charge of the binding pocket will ultimately determine what role E90 plays in ligand binding in the pocket, as well as the nature of such a role.

Alternate Binding Position for Ethidium

The E90Q-QacR-Et complex structure indicates that another position for ethidium binding is available to QacR beyond the position visualized in the WT-QacR-R6G complex structure. This new position for the ethidium in the pocket is the only position for ethidium observed in this structure. As the apparent increase of affinity in the new position in the pocket is only two fold it should not completely abrogate binding of ethidium in the original ethidium binding position even if the sites are mutually exclusive. With the measured difference in affinity, the ratio of the occupancy of ethidium at each position should be 3:2. Thus, ethidium should be visible at both positions. It is not clear why the new position for the ligand is preferred. The new binding position for the ligand does not have a significantly greater affinity compared to the previously observed position of ethidium, yet ethidium has not been observed in the new binding position before.

It is possible that E90 has secondary effects on the structure of the binding pocket. Observed changes in the orientations of the binding pocket residues E90, Y103, and E120 create an alternate binding site. Yet, it is not clear how the substitution of E90 would effect the orientation of these residues. Similarly F89 from the bovine odorant-binding protein (bOBP) is found in alternate binding positions for no visible structural reason (*84*). Most ligands will bind in both of the two subunits of this protein in very similar

binding positions. However, the depending on the drug binding the position of F89 is effected, but not symmetrically between the two subunits. For example, with undecanal bound to bOBP, F89 is in one position in the A subunit and in the other in the B subunit (*84*). It is believed that this asymmetry is produced with subunit crosstalk through the segments of the subunits that cross over (*84-86*). Similar to a possible secondary effect by E90 how this communication occurs is not clear.

The binding position of the ethidium may be influenced by the differences in the conformations of the E90Q-QacR-Et and WT-QacR-Et complexes. The DNA binding domains of the E90Q-QacR-Et complex structure are much closer together than in the DNA binding domains in the WT-QacR-Et complex structure. These may affect very small changes that effect the position of helix 6, which runs along the side of the binding pocket. These small changes in the binding pocket may change the constraints and geometry of the pocket enough that one position of the pocket has a higher affinity over the other although in solution the affinities are the same. However, an alignment of the regions of WT-QacR-Et and E120Q-QacR-Et that are involved in ligand binding yields a small RMSD change of 0.7 Å. The accessible volumes of the binding pockets of E90Q-QacR-Et and Wt-QacR-Et, as measured by the program VOIDOO, indicate a difference of only about 10 $Å^3$. Furthermore using an overlay of the WT-QacR-Et and E90Q-QacR-Et structures, no structural changes that would favor one position of ethidium or the other could be found.

R6G in Two Ligand Binding Positions

The E90Q-QacR-ligand complex structures and affinity measurements indicate that QacR has multiple binding modes with similar affinities. There is no change in

affinity for R6G by E90Q to account for the second R6G binding position observed in the E90Q-QacR-R6G complex structure. This second site is mutually exclusive with the first. Earlier studies of the interactions of ethidium and proflavin with QacR indicate that it can bind these two ligands simultaneously (*8*). However, the X-ray crystal structure and stoichiometry data suggest that QacR binds one R6G at a time. The most likely reason for electron density at both of these positions simultaneously is statistical disorder. In some asymmetric units, R6G is one position in some unit cells it is in the other. The crystal structure is an average of the structures of these two unit cells.

The existence of multiple binding positions for a single ligand to one binding site has precedent, as the human multidrug-activated transcription factor PXR complex binds the drug SR12813 at an affinity of 41 nM in three distinct orientations (*87*). Similar to QacR multiple orientations of ligand binding in PXR are available as the binding pocket is large with evenly spaced polar residues that contact the ligand in various positions. This allows for multiple binding sites, each having similar interactions with the ligand. In the E90Q-QacR-R6G complex structure, the charge-charge interaction of R6G in position 2 with E120 is similar to the interaction of R6G with E90 in position 1.

One could argue that position 2 is a lower affinity site for R6G and the affinity measurements used could not detect the weaker binding of R6G to position 2. Furthermore, one could argue that this position is visible only with crystallography because of the innately high concentrations of ligand and protein used in this method. However, the relatively even distribution of electron density between the two positions in the pocket does not suggest this. It suggests the sites have similar affinities. Given that these positions are mutually exclusive, if one had a much higher affinity R6G in this

position would dominate in the structure at any protein concentration. The alternate position would not be significantly occupied. In the case of E90Q-QacR-R6G structure, the density in position 2 is better, but position 1 is still strongly occupied. The affinities for the two positions cannot be significantly different.

It is not clear from the structure why this second site is available for R6G binding. There is not a significant change of orientation of Q90 compared to E90. The residues adjacent to Q90 do not change in orientation either. Thus, it is difficult to see how substituting this glutamate with a glutamine allows this added mode of binding. It is also interesting that position 2 for R6G is in the ethidium-binding pocket. It occupies a volume that overlaps with both the ethidium position in the WT-QacR structure and overlaps to an even greater degree with ethidium in E90Q-QacR. If the affinity has not changed with the substitution and there are no obvious changes in the structure, binding at position 2 should be available to the R6G in the wild type QacR. A R6G at position 2 is not observed in the WT structure. However, there is no obvious structural reason that would discriminate against R6G in the ethidium-binding pocket.

It is not clear exactly what the role of glutamate 90 is in the structure. It has no apparent role in ligand affinity and it is not clear exactly how it influences the binding positions adopted by ethidium and R6G. As there is no sequence homology for the binding pocket of TetR family members, there is no known information on the necessity of these glutamates other than what is presented here. *In vivo* analysis of the ability for this mutant to survive a challenge with ethidium bromide may further elucidate significance of this substitution for ligand binding.

Tyrosines, Ligand Selectivity, and Binding Modes

Although there is some reorientation of the residues when ligands adopt different positions, Y123 remains in the same position. Ligand complexes with WT-QacR indicate that Y123 is basically the gate between the two pockets and that depending on the ligand and which part of the pocket it resides Y123 would adopt a specific orientation (*1*). In the case of double drug binding, the position of the Y123 is dominated by the position of one of the two ligands in the binding pocket (*8*). The tyrosine and the ethidium move to accommodate the proflavin, which does not change position (*8*). Although the R6G is statistically disordered in the E90Q-QacR-R6G structure it is effectively in both binding pockets. The Y123 remains in the orientation seen in the WT-QacR-R6G complex. In this way the first position of R6G, which binds to the same portion of the pocket as proflavin, may dominate the orientation of Y123. The other possibility is that although position 2 R6G is in the ethidium-binding pocket its preferred binding is a mode that does not require reorientation of Y123. This could be the case as R6G and E120 are statistically disordered to accommodate the two mutually exclusive positions of R6G in the binding pocket. The position of one R6G does not affect the other. If Y123 preferred a different position for position 2 R6G, then it should be statistically disordered as well. Thus although this second position for R6G occupies a portion of the pocket closer to that of Et, Y123 maintains the same orientation seen with ligands bound in the R6G pocket.

The new positions of ethidium and R6G in these structures are more central to the binding pocket compared to previous ligands. They share more components of both pockets and may support a new binding mode. Although the ethidium and R6G from E90Q-QacR ligand structures overlap with the position of the ethidium from the WT-

QacR-ethidium complex they both interact with W61, a residue characteristic of the R6G binding pocket. Rhodamine-6G in position 2 also interacts with Y93, another contact considered a hallmark of binding to the R6G binding pocket. This intermediate position between the R6G and Et binding pockets is similar to the position of MG. In an alignment of the E90Q-QacR-Et and -R6G structures with WT-QacR-MG, the three ligands overlay remarkably well, especially R6G in position 2. Furthermore, Y123 is in the same position in the E90Q-QacR-Et, E90Q-QacR-R6G, and WT-QacR-MG structures. The ethidium and R6G from these structures could be considered to be a malachite green binding mode.

Other Small Molecules in the Binding Pocket

In these studies, not only was there new electron density for ligands discovered but there was new electron density for solvent components as well. Identification of imidazoles in the E90Q-QacR-MG and E90Q-QacR-Et structures was not surprising, as at the pH of the crystallization conditions, imidazoles fit the lipophilic cation profile. Furthermore, the concentration of imidazole in the crystallization experiment starts near 25 mM and is concentrated with the protein in the process of crystallization. As imidazole has been shown to inhibit QacR-ligand in our binding studies, imidazole must bind to QacR. The ligands, however, have a high enough affinity and high enough concentration to compete with the imidazole. But as MG binds closer to the surface of the protein part of the pocket is left open; it appears imidazole fills the leftover portion. Similar to this discovery, molecules of imidazole were discovered in the ligand-binding pocket of BmrR, another protein that binds lipophilic cations (*30*).

The good fit of bromide ions into the remaining electron density proximal to the ethidium in E90Q-QacR-Et suggests that the other ligands may be accompanied by counter ions. This suggests that the ethidium can bring its own counter ion into the pocket. This counter ion could make the net charge of the ethidium zero and eliminate its dependence on charge-neutralization by the protein. Thus, counter ions could eliminate the need of the negative charge of the glutamates for cationic ligand recognition. In the E90Q-QacR-Et complex structure the counter ions are good enough of an X-ray scatterer to be seen at this resolution. Acetate or chloride ions, which could be in the other ligand-complex structures, are not easily discernable at the resolutions of most of the structures. However, given the number of tyrosines and other aromatic residues that may be able to provide the partially negative surface of a π-system for a π-cation interaction with the ligand, inclusion of a counter ion may not be necessary.

Mechanism of Induction Revisited

The conformation of QacR observed in the E90Q-QacR-MG structure suggests that the large movements of the recognition helices observed in the WT-QacR-Dq complex structure may not be necessary for induction. The DNA-binding domains of the $P6_2$ conformation of the ligand bound form of QacR are much closer together. The distance between the recognition helices is the same as that of the WT-QacR-DNA structure. This suggests that it is not necessary that an increase in the distance of the recognition helices is required for induction. As the ligand-bound subunits of both the $P6_2$ and $P4_22_12$ space group structures have undergo similar twisting motions upon induction, rather than an increase of the distance of the DNA-recognition helices, the twisting motions of the DNA binding domains on ligand binding may dominate the

mechanism of induction. One caveat is that the distances between the C_α of Y40 of each DNA-binding domain in the E90Q-QacR-MG complex is 1.6 Å larger than in the WT-QacR-DNA complex. As the Y40 residues makes a contact with a base from the DNA, this increase in distance between the residues may be enough to affect the contact with DNA. For example, TetR shows only changes of 3.0 Å between its recognition helices and the affinity of this repressor for DNA is tightly regulated.

E90Q-QacR-MG structure is a snap shot of an alternate conformation of QacR that illustrates the flexibility inherent to the protein. Crystal structures of other TetR family members in the ligand bound state show similar conformational flexibility. In EthR, the distances between the DNA recognition helices of ligand bound crystal structures from two different space groups are 47.8 and 52.3 Å (*33; 41; 44*). In comparison, the distances between the DNA recognition helices are 40.9 and 42.4 Å in ligand bound CprB crystal structures from two different space groups (*32*). Similar to the difference between the DNA recognition in the $P6_2$ and $P4_22_12$ space group QacR crystals, the interhelical distances of these ligand bound TetR homologs range from 2 to 13 Å larger than the distance of the DNA recognition helices from the WT-QacR-DNA complex structure.

Summary

E90Q influences the positions of ligands in the binding pocket but it is not critical for their affinity. It is also not clear if there is a relationship between the negative charge of the glutamate residue and the positive charges of the ligands that interact with it. Further studies on other glutamates in the binding pocket are required to see if these residues are also not critical for ligand binding. The individual charges of the glutamates

may not be required as the binding pocket as a whole is highly electronegative and π-cation interactions with the aromatic residues may substitute for a formal negative charge of a glutamate. The WT-QacR-pentamidine structure illustrates this as in this structure the hydroxyl of a tyrosine can complement the positive charge of a ligand (*7*). It is also possible that the binding pocket may include counter ions for the ligands. Moreover, it is possible that E90 has a role that effects the orientation of the residues in the pocket by some yet unidentified mechanism.

Comparison of a crystal structure from the new $P6_2$ space group with previous crystal structures has yielded some interesting new insights on the mechanism of QacR induction and DNA binding. Analysis of the $P6_2$ E90Q-QacR-MG complex structure indicate that the angular component of the structural change induced on ligand binding may be more important to QacR induction mechanism than is the distance between the helices. Further analysis of the apo structure indicates that it has a significantly different conformation than the WT-QacR-DNA complex structure and the WT-QacR-ligand complex structures. This structure may yield insight on the mechanism of DNA binding by QacR.

The most significant finding provided by this research is that the QacR has multiple binding positions of similar affinity for a ligand. This further illustrates how QacR is an ideal multidrug binding protein. Not only can QacR bind numerous drugs with high affinity, it can also bind the same drug in different ways with the same affinity. The degenerate ligand binding further illustrates the promiscuity of the QacR ligand-binding pocket.

Figure 1. ITC binding data for E90Q-QacR and Dq. The top panel is the raw multiple injection ITC binding data and the lower panel is the binding isotherm determined from that data. The average values for the experiments are included.

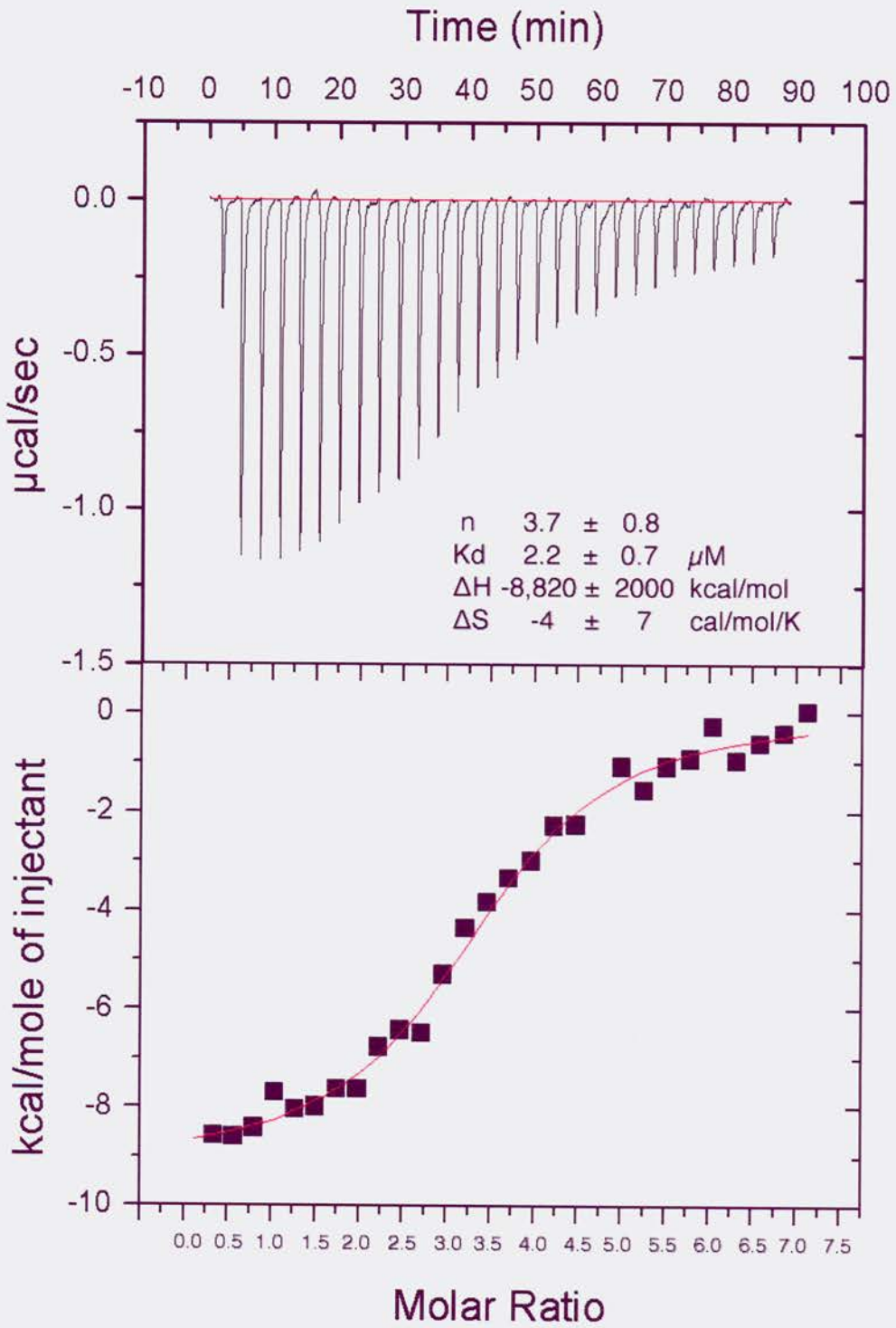

Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100
µcal/sec
0.0
-0.5
-1.0
-1.5
n 3.7 ± 0.8
Kd 2.2 ± 0.7 µM
ΔH -8,820 ± 2000 kcal/mol
ΔS -4 ± 7 cal/mol/K
kcal/mole of injectant
0
-2
-4
-6
-8
-10
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5
Molar Ratio

Figure 2. QacR ligand complex crystals. A. Panel A shows an E90Q-QacR-MG complex crystal with the morphology typical of the $P6_2$ space group crystals. B. Panel B shows the $P6_2$ space group E90Q(K67S)-QacR-R6G complex crystal, which did not usually grow larger than 100 μm. C. Panel C depicts examples of $P4_22_12$ space group E90Q(K67S)-QacR-R6G complex crystals. Some of these are epitaxially twinned. D. Panel D shows an example of a E90Q(K67S)-QacR-Dq crystals which can grow as large as 700 μm but are highly twinned and diffract poorly. E. Panel E shows a few examples of the P62 space group E90Q(K67S)-QacR-Et crystals. Though these crystals have an odd morphology and were obviously twinned usable data was able to be collected.

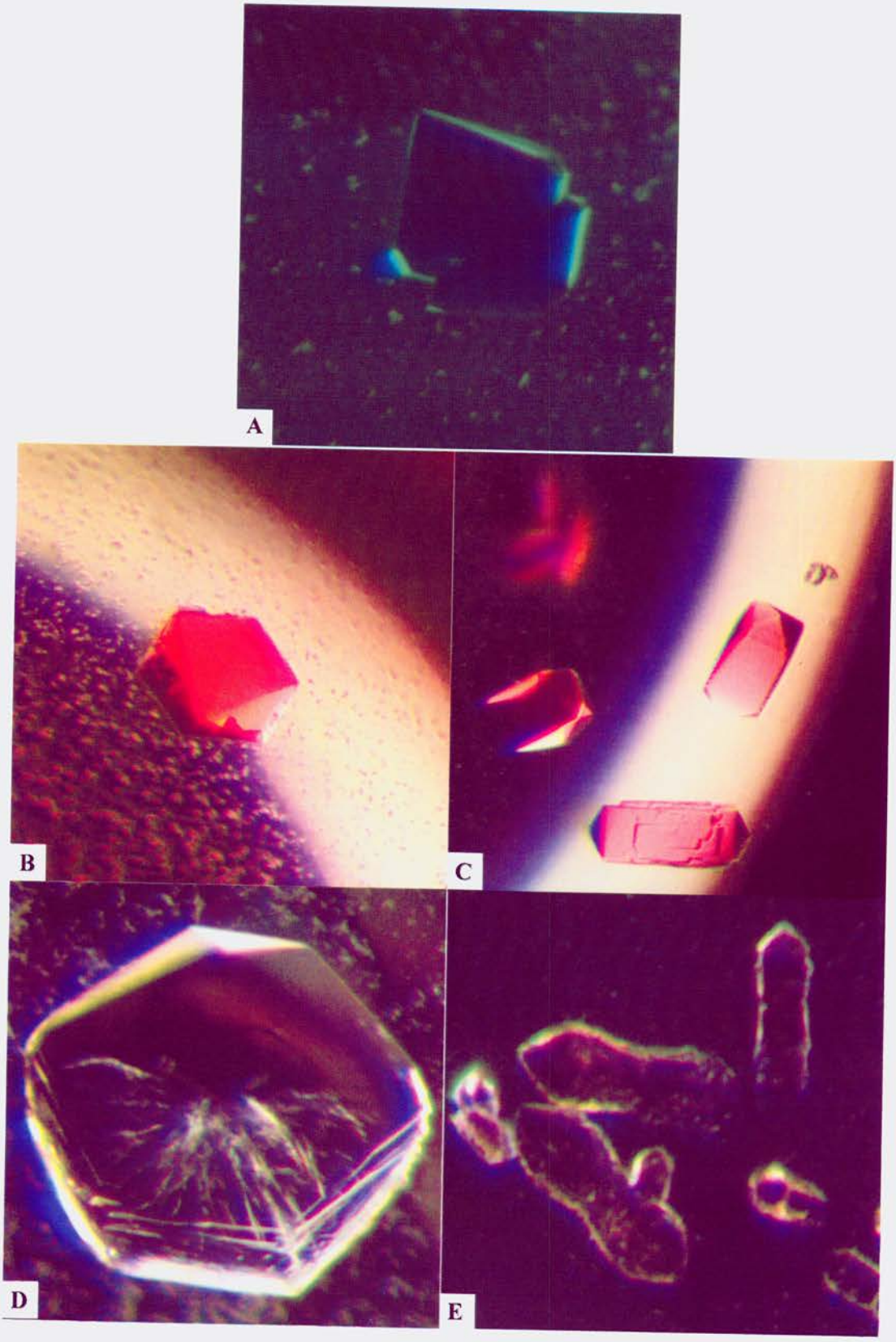

A
B
C
D
E

Figure 3. Binding pockets of the E90Q-QacR-MG and WT-QacR-MG complexes. The E90Q-QacR-MG complex (blue) and WT-QacR-MG complex (yellow) structures are identical except for the imidazoles that appear in the E90Q-QacR-MG complex structure.

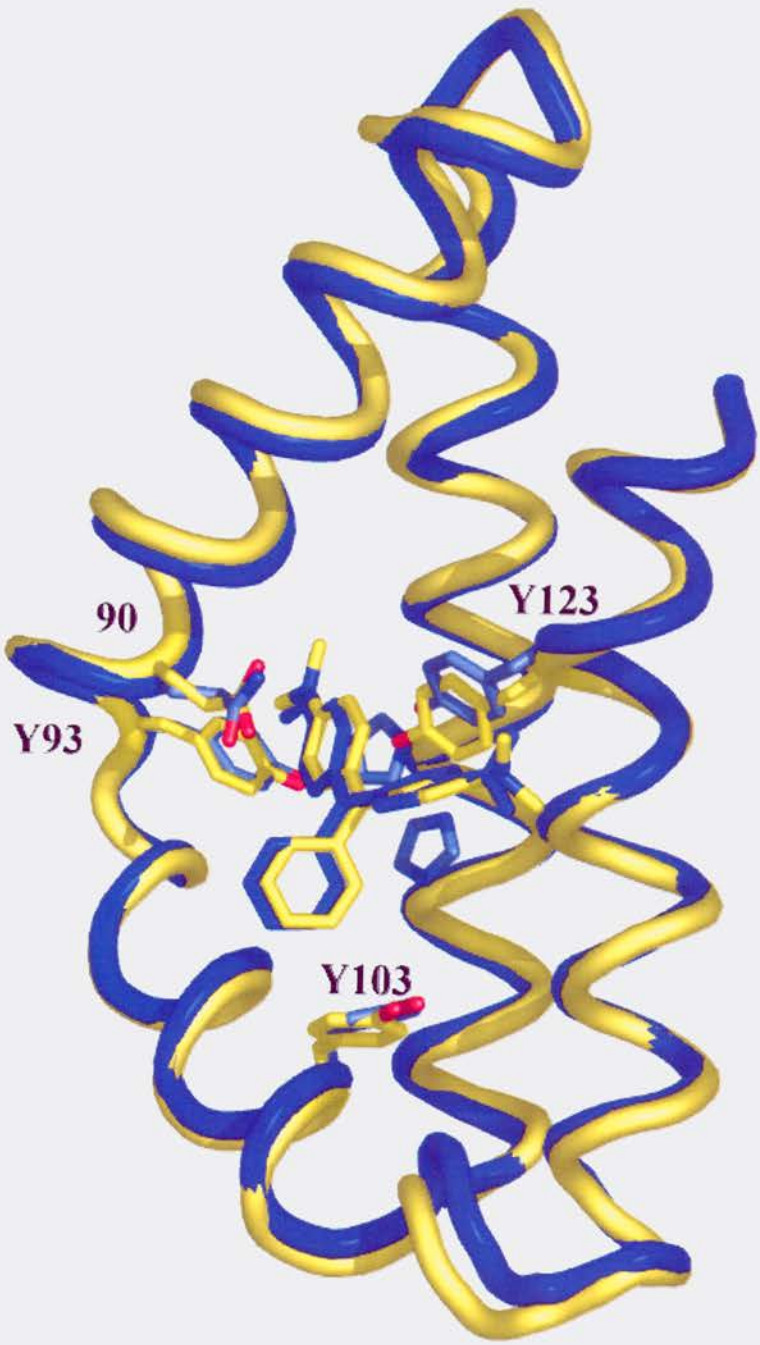

90
Y123
Y93
Y103

Figure 4. Overlay of the WT-QacR-Dq and E90Q-QacR-MG complex structures. This overlay compares the over all conformation of QacR from the WT-QacR-Dq complex structure (yellow) from $P4_22_12$ space group crystals and the conformation of QacR from the E90Q-QacR-MG complex structure (blue) from the $P6_2$ space group crystals. The structures overlay well except for the DNA binding domains. The DNA binding domains are closer together in the E90Q-QacR-MG complex structure.

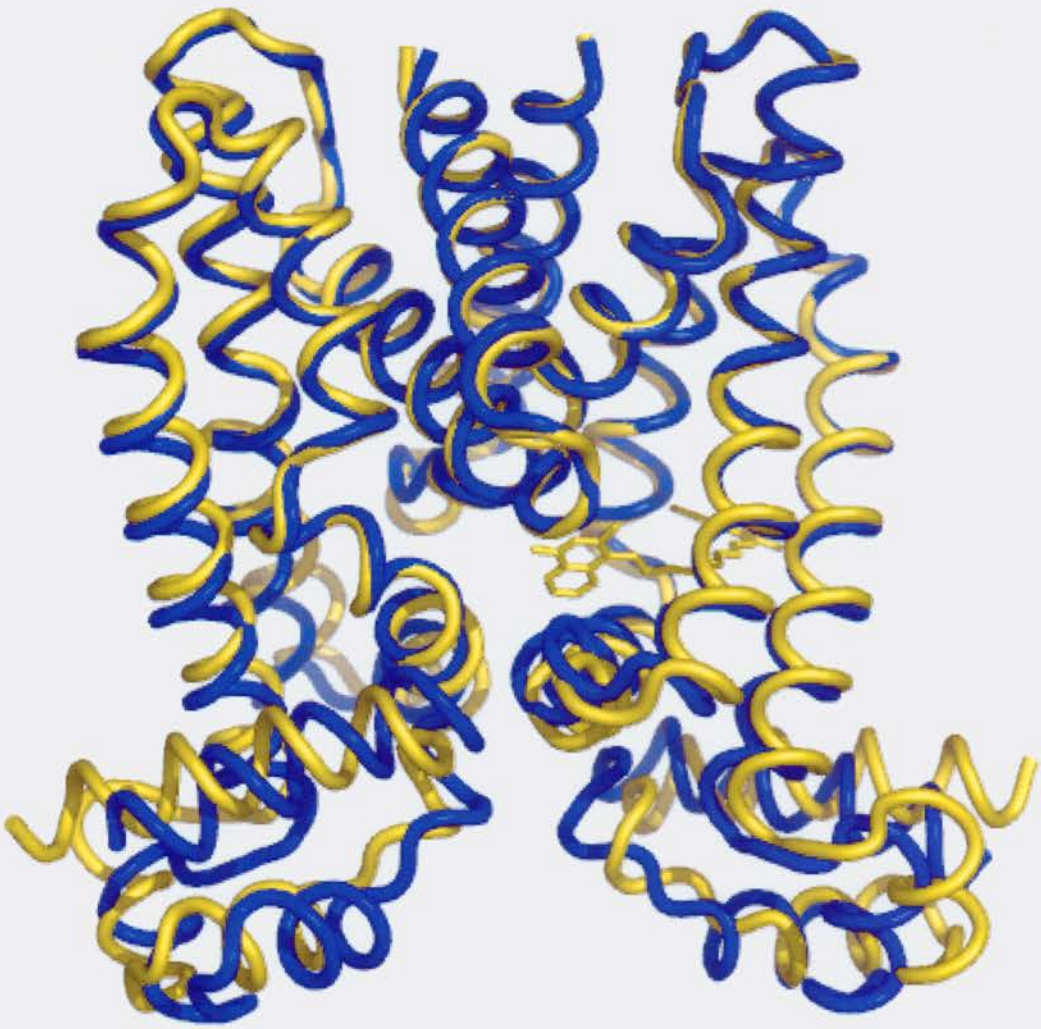


Figure 5. DNA binding domains from the two conformations of QacR. The E90Q-QacR-MG complex structure (blue and purple) and WT-QacR-Dq complex structure (yellow and orange) were superimposed on the B subunit. The DNA recognition helices from the WT-QacR-Dq complex structure are 6.6 Å farther apart than in the E90Q-QacR-MG complex structure.

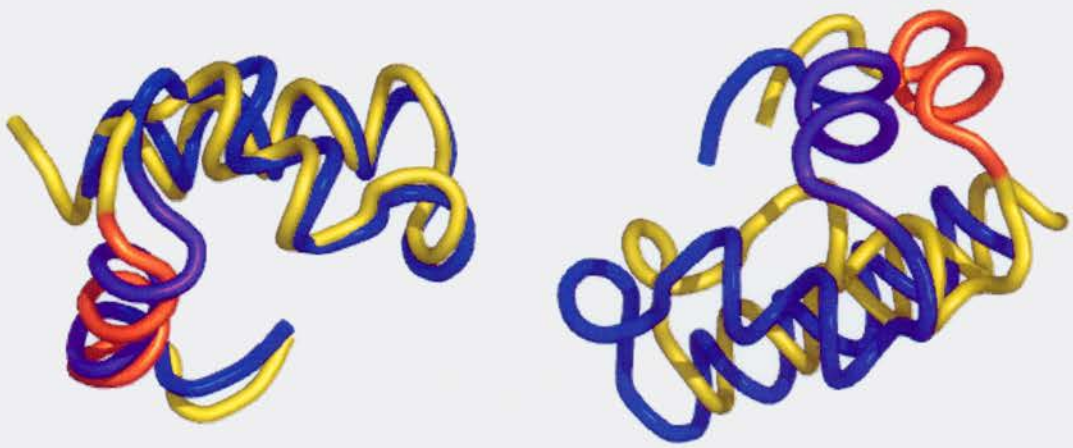
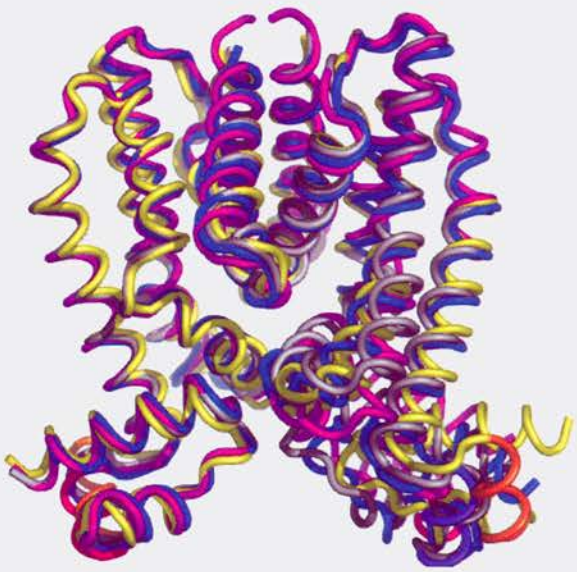


Figure 6. Angular displacement of the DNA recognition helices. A. QacR in four structural states are overlaid by alignment with the B subunit. These states include two different drug bound states, E90Q-QacR-MG complex (blue and purple) and WT-QacR-Dq complex (yellow and orange), as well as the WT-QacR-DNA complex (magenta and pink) and the WT-QacR-Apo complex (grey and black). B. The DNA binding domains from the A subunits of the E90Q-QacR-MG complex, WT-QacR-Dq complex, and the WT-QacR-Apo complex cluster together. C. The angles of the DNA recognition helices relative to the orientation of the DNA recognition helices from the WT-QacR-DNA complex in the overlay are 21° for the E90Q-QacR-MG complex and WT-QacR-Apo complex and 26° for the WT-QacR-Dq complex.

A



B

C

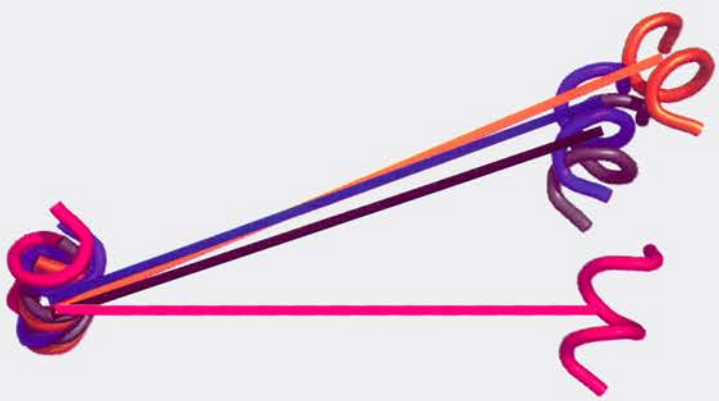
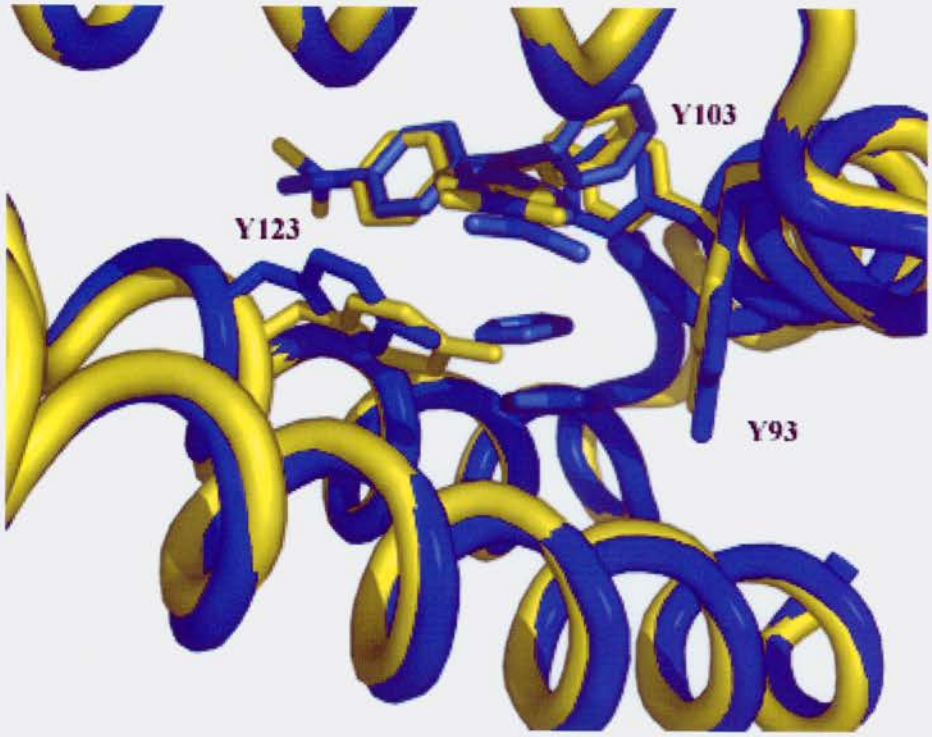
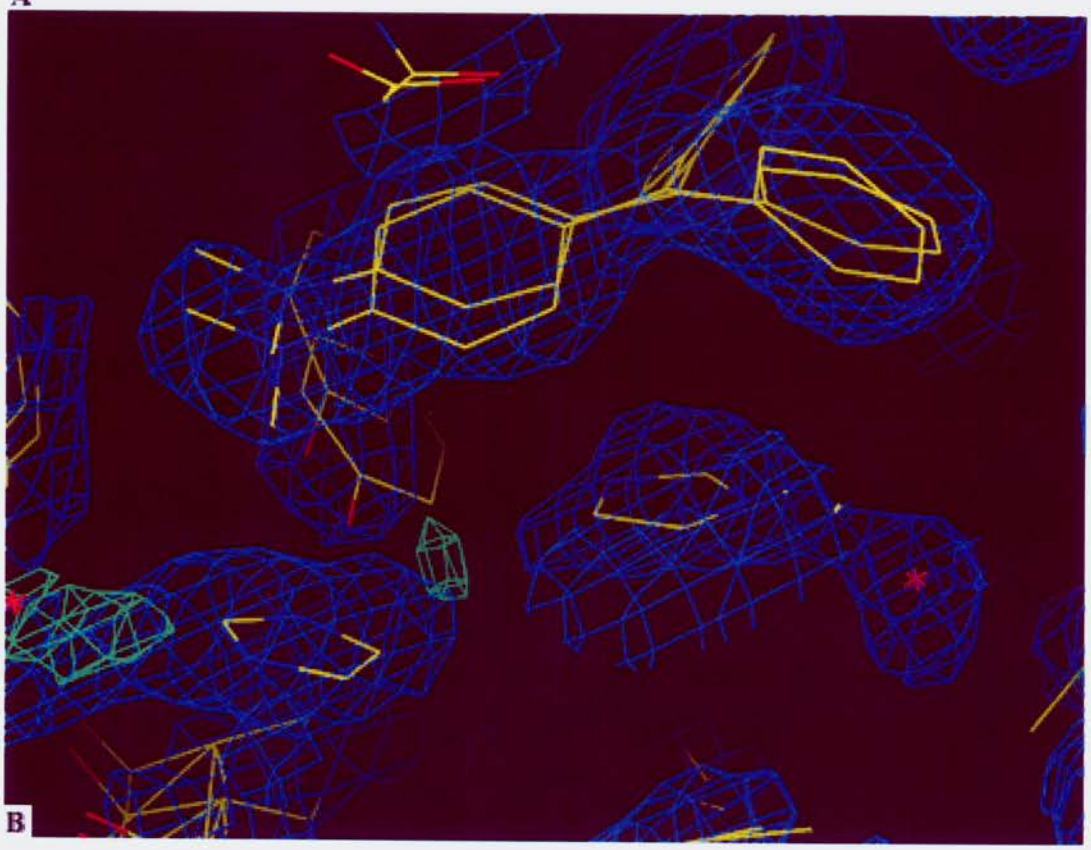


Figure 7. Imidazoles with MG in the binding pocket. A. As malachite green does not bind deeply in the binding pocket a large aromatic pocket remains and is occupied by imidazoles. The structures in this figure are the WT-QacR-Dq complex (yellow) and the E90Q-QacR-Dq complex (blue). B. Imidazole and water in $2F_o$-F_c (blue) composite-omit-density contoured at 1 σ and the remaining F_o-F_c (green) electron density contoured at 3 σ.



A



B

Figure 8. Overlay of the binding pockets of the WT-QacR-Dq and E90Q-QacR-Dq complexes. A. Compared to the position of Dq in the WT-QacR-Dq complex (yellow), dequalinium moves slightly in the E90Q-QacR-Dq complex (blue) structure accompanied by a change in the position of E90Q and H164. B. The $2F_o$-F_c (blue) composite-omit-density (contoured at 1 σ) indicates that the fit of the E90Q-QacR-Dq complex model is good.

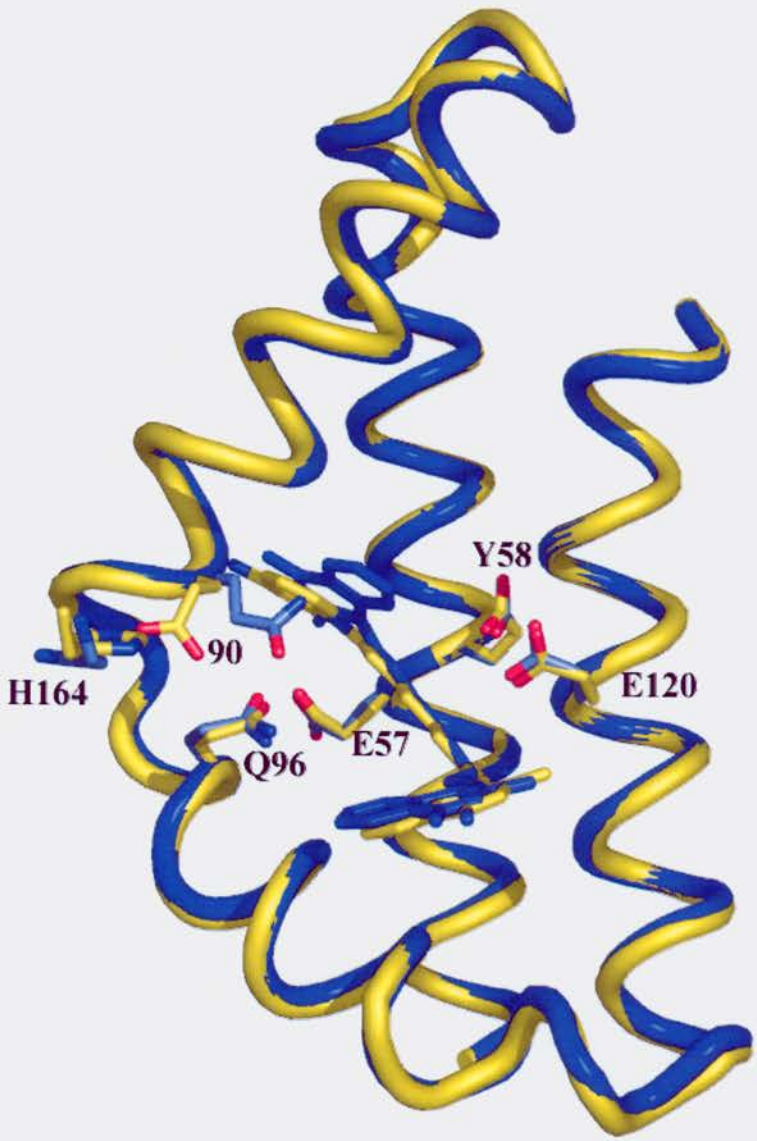

Y58
90
H164
E120
E57
Q96

Figure 9. Overlay of the binding pockets of the E90Q-QacR-Et and WT-QacR-Et complexes illustrating changes in residue orientation. Comparing the E90Q-QacR-Et (blue) and WT-QacR-Et (yellow) complex structures, the ethidium flips over and moves towards tryptophan 61 in the E90Q-QacR-Et complex structure. Residues E90Q, E120 and Y103 reposition themselves to accommodate the new position for ethidium.

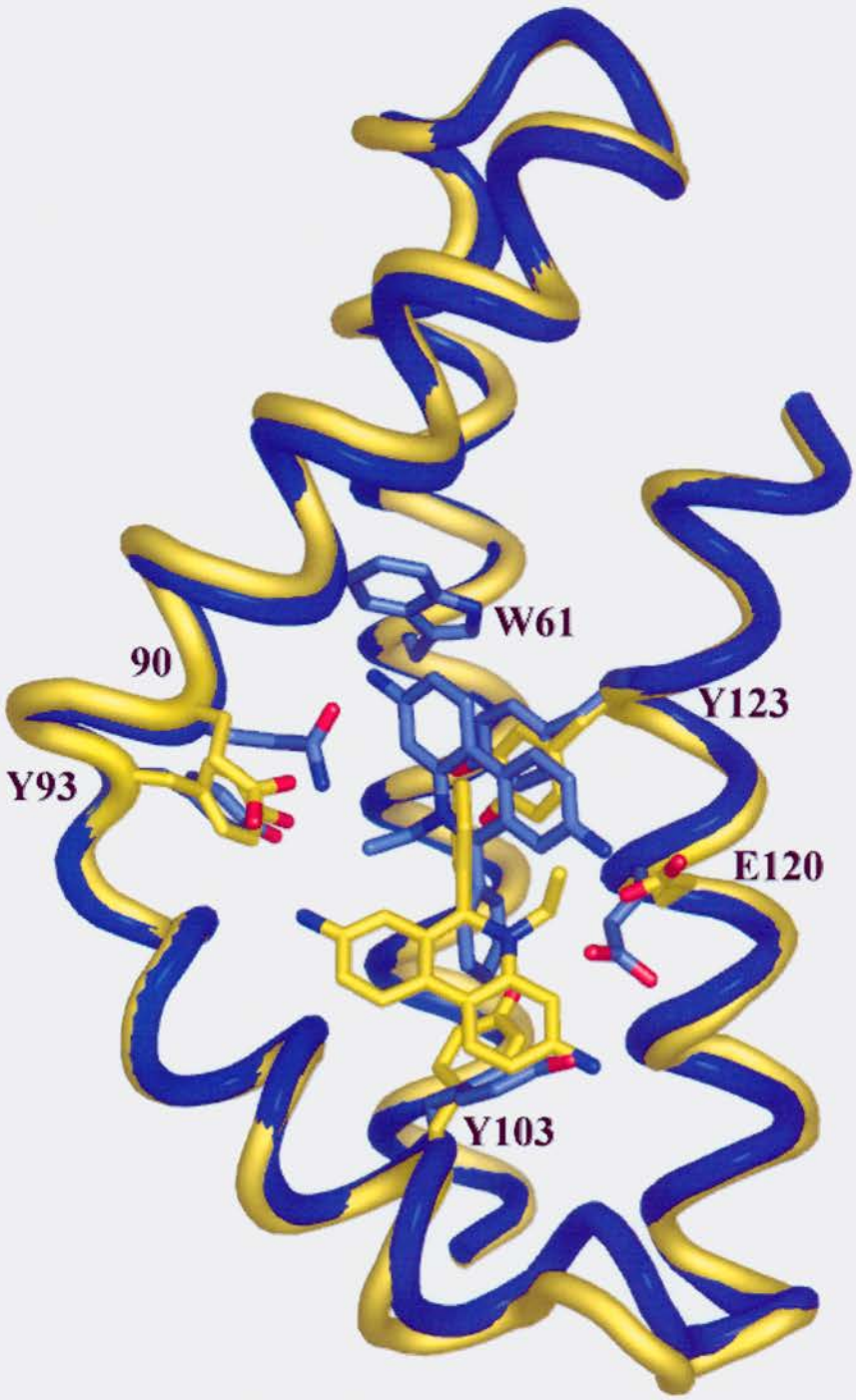

W61
90
Y123
Y93
E120
Y103

Figure 10. Overlay of the binding pockets of WT-QacR-Et and E90Q-QacR-Et complexes indicating new ligand residue interactions. Comparison of the position of ethidium from the WT-QacR-Et (yellow) and E90Q-QacR-Et (blue) complex crystal structures indicate make new contacts with the ligand.

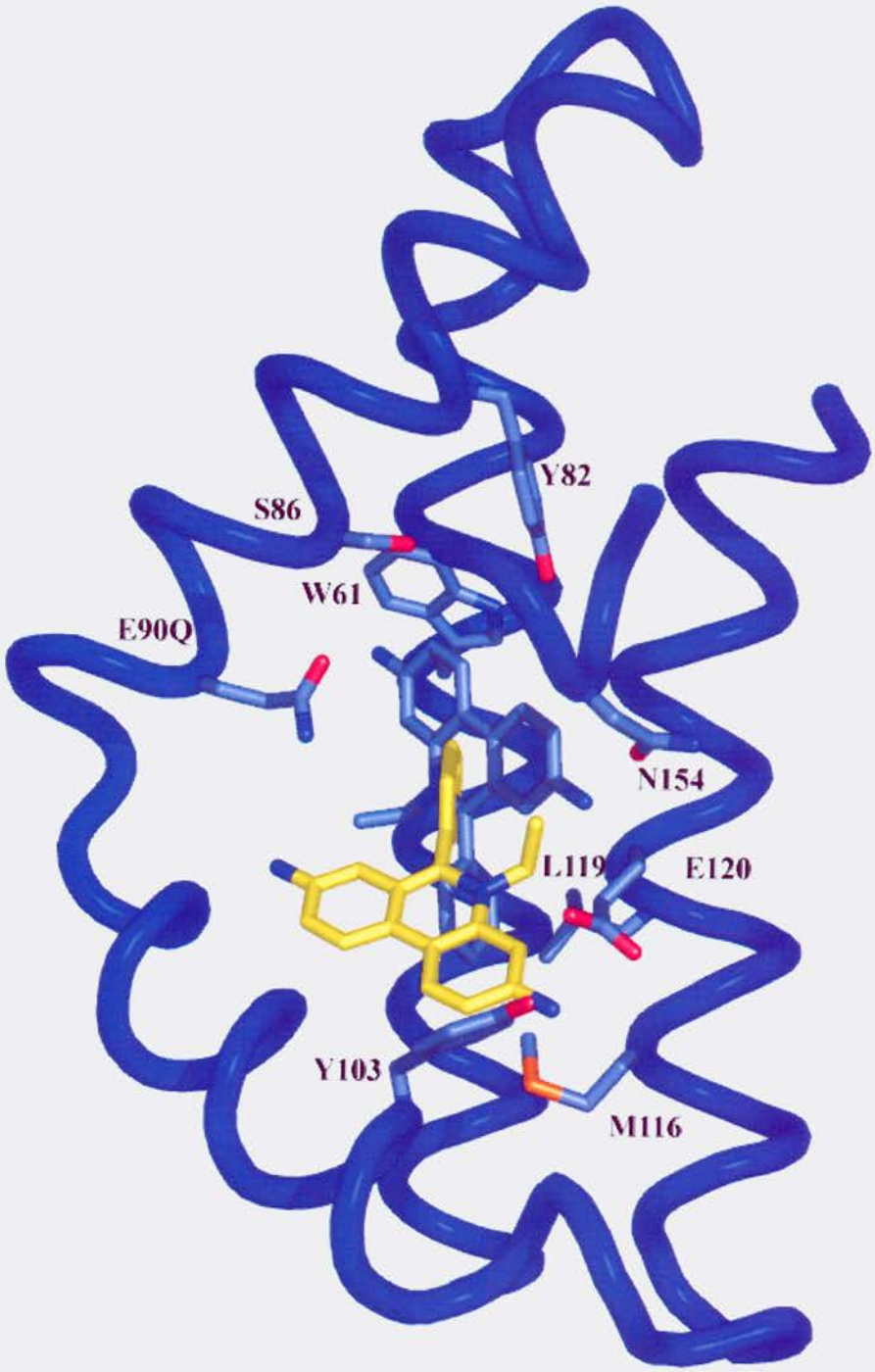

Y82
S86
W61
E90Q
N154
L119
E120
Y103
M116

Figure 11. Electron density maps of the E90Q-QacR-R6G complex binding pocket. Two large areas of F_o-F_c map (contoured at 3 σ) and $2F_o$-F_c composite-omit-map (contoured at 1 σ) electron density correspond to two overlapping molecules of R6G.

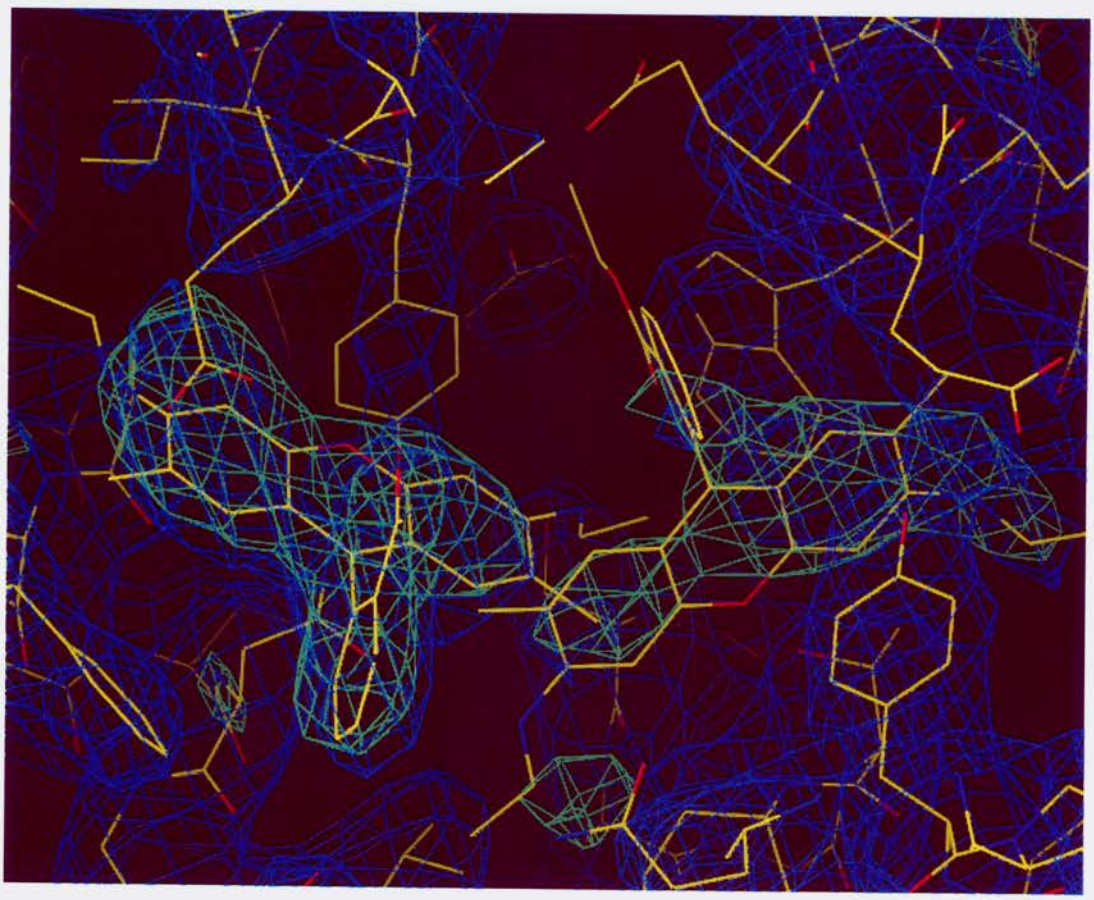
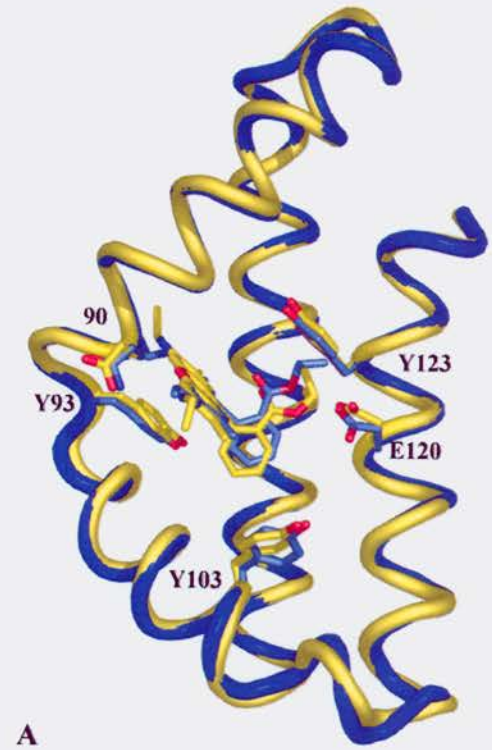
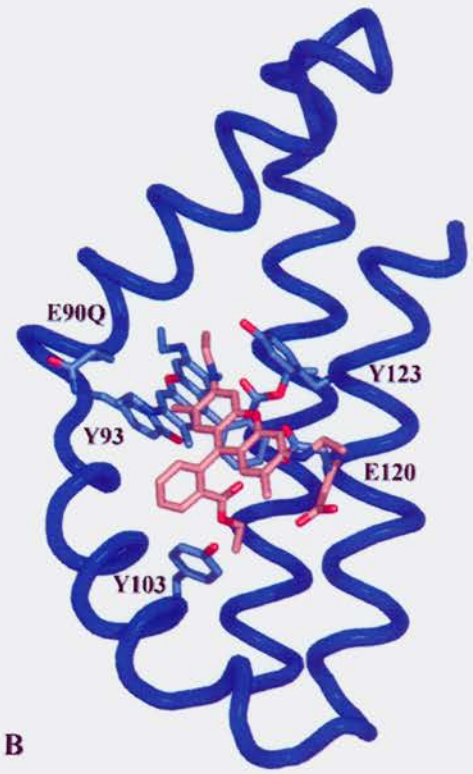


Figure 12. Position 1 and position 2 of R6G binding. A. Position 1 (blue) for R6G binding is identical to R6G binding in WT-QacR-R6G (yellow). B. Position 2 (pink) for R6G binding is in the ethidium-binding pocket and E120 moves to accommodate the ligand.



A



B

Figure 13. Fluorescence polarization based stoichiometry measurements of R6G for E90Q-QacR.

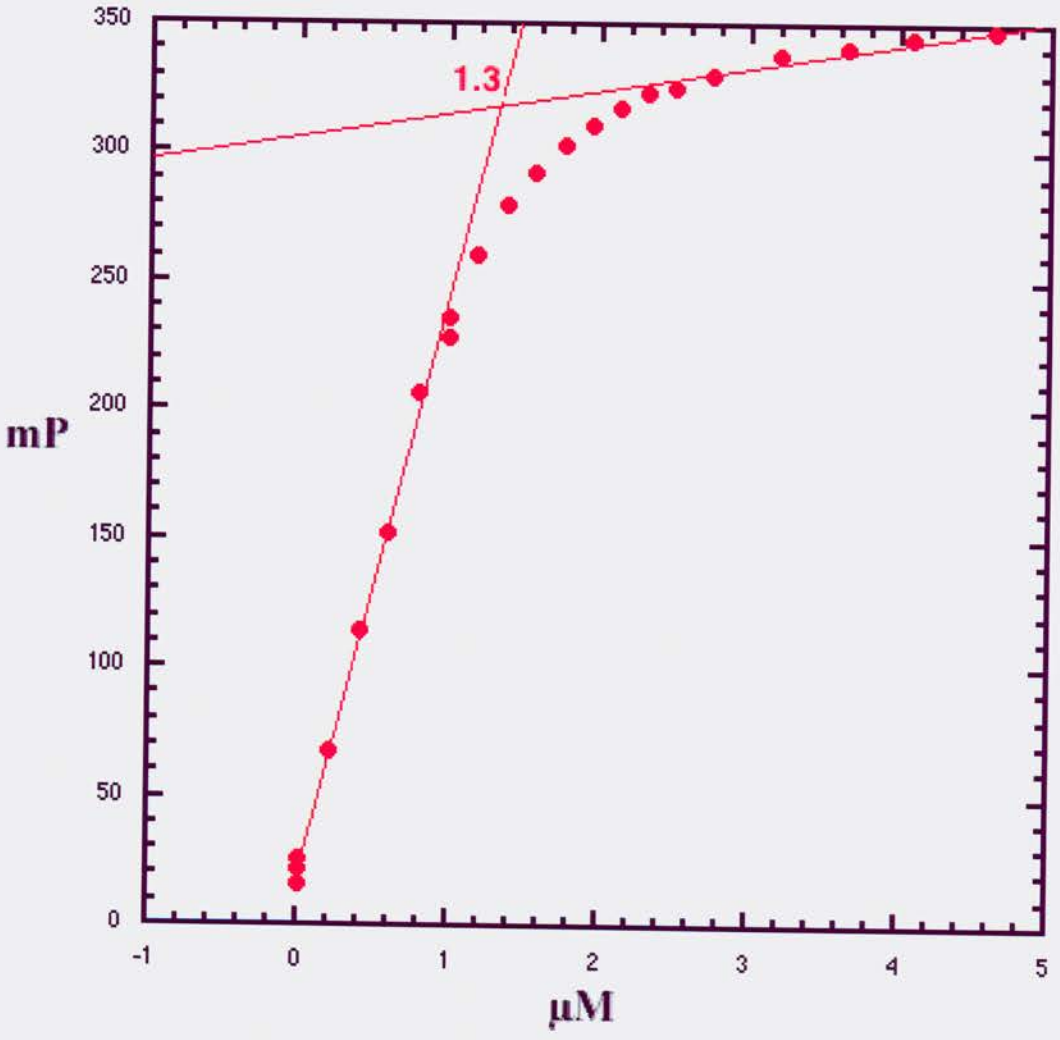
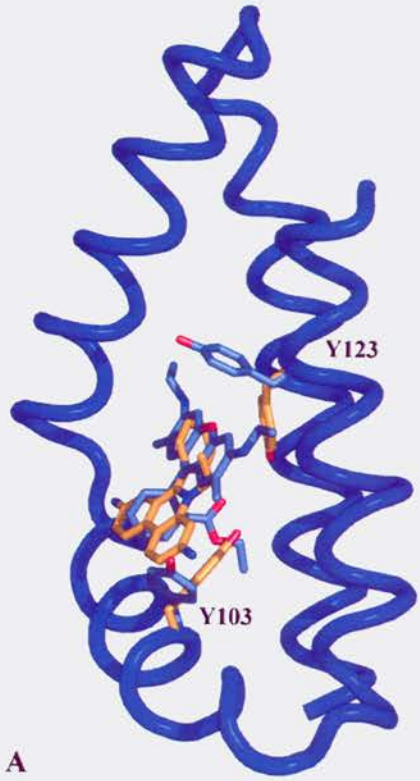
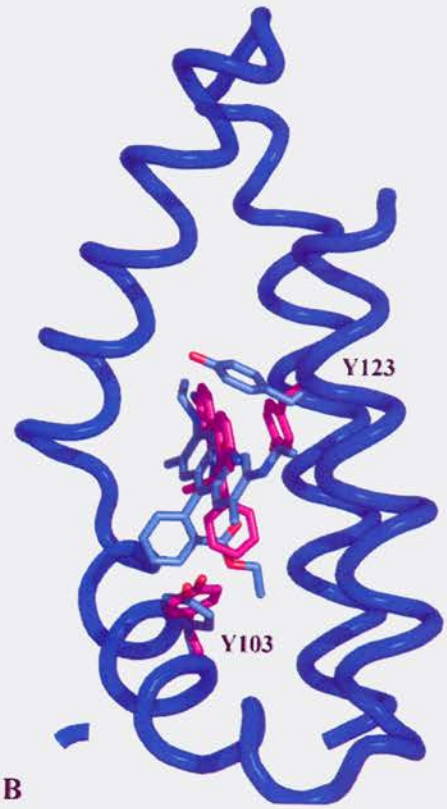

350
300
250
200
150
100
50
0
mP
1.3
-1
0
1
2
3
4
5
µM

Figure 14. Position 2 of R6G compared to ethidium binding. A. The R6G (blue) from E90Q-QacR-R6G binds the same site as ethidium (orange) from WT-QacR-Et but it is flipped over by 180°. Tyrosine 103 and 123 are both in different orientations in these structures. B. The R6G from E90Q-QacR-R6G binds even more similarly to ethidium (pink) from E90Q-QacR-Et. The orientation of tyrosine Y123 is different between these two structures.



A



B

Table 1. E90Q-QacR ligand dissociation constants.

	WT	E90Q
ET	2.4 ± 0.1 μM	1.18 ± 0.04 μM
MG	1.2 ± 0.1 μM	1.74 ± 0.02 μM
R6G	0.39 ± 0.02 μM	0.48 ± 0.02 μM
Dq	1.0 ± 0.5 μM	2.2 ± 0.6 μM

Table 2. X-ray diffraction and model statistics for E90Q-QacR-ligand structures.

		E90Q-QacR-MG	E90Q-QacR-Dq	E90Q-QacR-Et (K67S)	E90Q-QacR-R6G (K67S)
Temperature (K)		100	100	100	100
Space Group		$P6_2$	$P4_22_12$	$P6_2$	$P4_22_12$
Cell Constants (Å)		104.9	170.2	104.2	171.8
		104.9	170.2	104.2	171.8
		93.2	93.6	98.6	94.8
Resolution Range (Å)	Low	65.9	82.7	43.2	49
	High	2.2	3.3	2.8	2.9
CV-Luzzati Coord. Error		0.4	0.5	0.5	0.5
Overall Rsym (%)		6.2	7.9	6.1	5.9
Completeness		99.8	100	100	99.6
Overall I/sig(I)		7.4	8.9	18.1	14.2
Total Reflections		207700	148090	113089	126444
Unique reflections		31776	21098	15047	31642
Multiplicity		6.5	7	7.45	4
High Res. Shell	Res.(Å)	2.3-2.2	3.5-3.3	2.9-2.8	3.0-2.9
	Rsym(%)	42.3	38.6	40	47.9
	I/Sig(I)	1.8	2.0	4.6	2.8
	R_{work} (%)	22.4	21.4	20.8	22.6
	R_{free} (%)	25.7	28.9	26.9	29.2
RMSD	Bond Angles	1.1	1.4	1.3	1.4
	Bond Lengths	0.010	0.010	0.011	0.010
	Main Chain B Values	2.7	2.8	2.7	2.4
Solvent	Water	92	31	33	27
	Sulfate	3	17	9	15
Ramachandran Analysis					
Most Favored (%)		90.8	81.3	89.9	82.4
Additional Allowed (%)		7.2	16.3	8.6	15.7
Generously Allowed (%)		1.1	1.6	0.3	1.0
Disallowed (%)		0.9	0.9	1.1	0.9

Table 3. Table of Contact Distances for Imidazoles in E90Q-QacR-MG.

	Ligand Atom	Partner Atom	E90Q (Å)
IMD 401	IMD N1	Y123 OH	2.6
		E58 OE1	3.3
	IMD C2	TRP61 CB	3.6
		E58 N	3.9
		E57 C	3.8
		E57O	3.7
	IMD N3	Y93 OH	3.1
	IMD C4	Y93 CE1	3.5
		MG C23	3.9
	IMD C5	IMD C4	3.7
IMD 402	IMD N3	Y123 OH	4.2
	IMD C2	L119 CB	4.1
		E120 CG	3.7
	IMD N1	MG C3	4.0
		E120 OE2	5.3
	IMD C4	MG C8	4.2
	IMD C4	IMD C5	3.7

Table 4. Comparison of Contact Distances for Dequalinium in E90Q-QacR-Dq and WT-QacR-Dq.

Ligand Atom	Partner Atom	E90Q (Å)	WT (Å)
C1	Y103 CB	3.8	3.6
C2	I100 CA	3.7	4.3
	I100 CD1	3.8	4.2
C3	I100' CG2	3.9	4.8
C4	I100' CD1	3.5	4.1
	N97 OD1	3.7	3.5
C5	Y103 CG	3.6	3.5
C6	Y103 CD2	3.5	3.5
	F162' CE1	3.8	3.7
C7	Y103 CE2	3.4	3.5
C8	Y103 CZ	3.5	3.4
	F162' CD1	3.6	3.4
C9	Y103 CE1	3.6	3.4
	F162' CD1	3.7	4.0
N1	Y103 CD1	3.7	3.4
	F162' CE1	4.1	5.0
N2	W61 CG	4.0	4.3
N3	T89 OG1	2.2	4.0
C11	W61 CD	4.0	5.7
C12	Y93 CE2	3.7	3.5
C13	Y93 CD2	3.9	3.7
C13	W61 CE3	3.8	4.1
C14	W61 CZ3	3.8	4.1
C15	S86 O	4.1	3.4
C16	Y123 CE2	3.2	8.7
C17	Y123 CE1	4.1	6.5
C19	E57 CB	3.6	3.5
	Y93 OH	3.7	3.4
C23	L54 CG	3.5	3.8
	L119 CD1	3.7	4.8
	L119 CG	4.2	4.4

Table 5a. Comparison of Contact Distances for Ethidium in E90Q-QacR-Et and WT-QacR-Et.

	E90Q		WT	
Ligand Atom	**Partner Atom**	**(Å)**	**Partner Atom**	**(Å)**
C1			F162' CE1	4.1
			Y103 CD2	3.3
C2			F162' CD1	4.1
			Y103 CD2	3.5
C3	Q90 OE1	3.3	Y103 CE2	3.9
	F162' CD1	4.1		
C4	Y123 CE2	3.6	Y103 CE2	3.4
	Q90 CD	3.7		
	Q90 NE2	4.0		
N5	Y123 CE1	4.1	Y103 CE2	3.7
			Y103 CZ	3.9
			Y103 OH	3.9
C6			Y103 OH	4.1
C7	N157 OD1	3.2	I99 CG2	4.0
C8			I99 CG2	3.7
			I100 CG1	4.2
C9	N157 CB	4.0	I100 CG2	3.4
C10	Y123 CB	3.9		
C12	A157 ND2	4.1	F162' CE1	4.1
			Y103 CG	4.2
C13	Y123 CG	3.7	Y103 CE2	3.3
			Y103 CD2	3.5
			F162 CD1	3.7
C14			F162 CD1	3.7
			Y103 CE2	3.3
C16			N157 OD1	3.5
			N157 ND2	2.9
C17	I99 CG2	4.2	N157 ND2	3.1
	Y103 CG	4.1		
	Y103 CD1	3.6		
C18	E120 OE2	4.3	Y123 CD1	3.8
			Y123 CE1	4.0
			Y123 CZ	4.1
			Y123 CE2	4.0
			Y123 CD2	3.8
			Y123 CG	3.7
C19	E120 CG	3.9	Y123 CE2	3.4
			Y123 CZ	3.5

(Continued on following Page)

Table 5b. Comparison of Contact Distances for Ethidium in E90Q-QacR-Et and WT-QacR-Et.

	E90Q		WT	
Ligand Atom	**Partner Atom**	**(Å)**	**Partner Atom**	**(Å)**
C20	E120 CG	3.9		
C21			E120 OE2	2.5
			Y123 OH	4.0
C22	Q96 NE2	3.6	N157 ND2	4.1
	Q96 OE1	3.9		
			E120 OE2	2.7
	Q90 NE2	3.6	E120 OE1	3.7
N23	Q90 OE2	3.5	T161' O	4.0
	W61 CZ3	3.7	F162' CA	4.0
	W61 CH2	3.7	F162' C	4.0
	W61 CE3	3.8	F162' C	3.9
	W61 CZ2	3.8		
	W61 CD2	3.9		
	W61 CE2	3.9		
N24	N154 OD1	3.3	E90 OE1	3.7
	E120 OE2	3.5	E90 OE2	4.1
			Q96 OE1	2.9

Table 6a. Comparison of Contact Distances for Rhodmine-6G in E90Q-QacR-R6G and WT-QacR-R6G.

	Ligand Atom	Partner Atom	E90Q (Å)	WT (Å)
RHQ201	C3	Y93 OH	3.8	3.7
	C4	Y93 CE2	4.1	3.4
		W61 CB	3.7	4.2
	C5	W61 CE3	3.9	3.5
	N2	Q64 NE2 (OE1)	3.5	3.9
	C6	T89 OG1	4.0	3.3
		T89 CB	4.2	3.6
	C7	RHQ202 C21	1.7	
	C8	RHQ202 C21	1.9	
	C10	RHQ202 C21	1.3	
	C11	RHQ202 C21	1.0	
	C12	Q96 OE1	3.6	3.1
	C12	RHQ202 C21	1.3	
		RHQ202 C4	1.4	
		RHQ202 C3	1.2	
	C13	RHQ202 C21	1.7	
	C16	L54 CG	3.5	3.4
	C17	L54 CB	4.2	3.5
		L119 CD1	4.0	4.6
	C20	RHQ202 C3	1.0	
	N1	Q96 OE1	3.3	3.4
		Q96 NE2	3.5	3.2
	C21	W61 N	3.5	5.0
		K60 CB	4.1	5.3
	C22	Q90 CB	3.3	5.0
		Q90 NE2	3.2	OE2 4.6
	C23	T161 OG1	3.0	3.0
		V160 CG1	3.3	5.0
		Q90 NE2	3.3	5.0
	O27	RHQ202 N2	2.8	
	O2	E58 OE1	3.1	3.9
	C28	Y123 CD1	3.3	4.1
	C29	Y123 CA	3.1	3.5
		E58 OE1	2.9	3.3
		E58 OE2	3.0	5.5

Table 6b. Comparison of Contact Distances for Rhodmine-6G in E90Q-QacR-R6G and WT-QacR-R6G.

	Ligand Atom	Partner Atom	E90Q (Å)
RHQ202	O1	Y123 CD2	3.3
	C1	N157 ND2	3.4
	C3	N157 OD1	3.8
	C6	Y123 CD1	3.8
	C17	Y103 CD2	3.4
	C18	I99 CG2	3.3
	C20	E120 OE1	3.3
	C21	Q96 OE1	3.1
	C22	I124 CD1	3.6
	C24	Y123 CE1	3.9
	C25	W61 CH2	3.4
	C29	L54 CD2	3.4

Chapter 5: X-ray Crystallography Reveals a New Binding Mode for Ethidium Bromide with a Glutamate to Glutamine Substituted QacR at Position 120

Introduction

Glutamate 120, as one of the four glutamates in the large QacR ligand-binding pocket, appears to be critical for binding ethidium, Dq, and MG. Crystal structures of WT-QacR-ligand complexes indicate that the E120 comes is proximal to all of these ligands (*1*). Like all QacR ligands, these molecules are cationic, thus the charged nature of E120 may be essential for interactions with these ligands. To determine the degree to which the formal negative charge aids in ligand binding, we substituted glutamine at this position and studied the interactions of these substituted proteins for the structurally characterized QacR ligands: ethidium, MG, Dq, and R6G. The affinities of the substituted QacRs were measured with tryptophan fluorescence quenching, fluorescence polarization, and isothermal titration calorimetry. Interactions of these ligands with the E120Q-substituted QacR (E120Q-QacR) were also studied structurally with X-ray crystallography.

The data indicate that the presence of glutamate 120 is not crucial for ligand affinity. Similar to our observations with the E90Q-QacR-ligand complexes, none of the affinities were adversely affected by substitutions at this position; moreover as seen in the E90Q-QacR-Et complex crystal structure, ethidium bromide occupies multiple positions in the E120Q-QacR-Et complex crystal structures, suggesting that there may be multiple sites in the binding pocket with similar affinities for ethidium.

Results

Ligand Affinity

Substitution of E120 with a residue with no formal charge does not greatly affect ligand binding (Table 1). The affinity for DNA was that of wild type, which illustrated that the mutant protein was intact and functional. The binding affinities for E120Q-QacR for MG and R6G remain unchanged from the unsubstituted QacR. The binding affinity of E90Q-QacR for ethidium is two-fold greater than that of WT-QacR (K_ds of 1.18 ± 0.04 μM to 2.4 ± 0.1 μM). The binding affinity of E90Q-QacR for Dq is two-fold less that of WT-QacR (K_ds of 2.3 ± 0.5 μM to 1.0 ± 0.5 μM). However, the difference in the Dq affinity is close to the error of the measurements so this does not appear to be significant (Figure 1). The stoichiometry varied and it was not clear why it was not equal to 1.

Crystallization

The ease and characteristics of crystallization of the E120Q-QacR protein depended significantly on the ligand cocrystallized with it. Unlike the E90Q-QacR-MG crystals, E120Q-QacR-MG was a difficult complex to crystallize; yet it yielded the highest resolution data among all the E120Q-QacR-ligand crystals. Like E90Q-QacR-R6G and E90Q-QacR-Et, E120Q-QacR required the K67S substitution in order to cocrystallize with MG. At a resolution of 2.4 Å, these crystals diffracted better than any of the wild-type crystals and are second only to the E90-QacR-MG complex crystals in resolution. Similar to the E90-QacR-MG complex crystals, these E120Q-QacR-MG

complex crystals formed hexagonal ($P6_2$) crystals exclusively. However, unlike the E90Q-QacR-MG crystals, these crystals were significantly twinned (41-48%), and required data collection with multiple crystals before the amount of partial twinning of the crystal was low enough (30%) to be detwinned with the program CNS. The E120Q-QacR-R6G and E120Q-QacR-Et complexes also required the K67S substitution in order to crystallize. However, like the E90Q(K67S)-QacR-Dq crystals, the E120Q(K67S)-QacR-Dq complex crystals form large crystals easily, but these crystals did not diffract well, were epitaxially twinned, and were useless for data collection.

E120Q-QacR-Et complexes did not form crystals *de novo* as the other complexes had. Seeding with small, whole E120Q-QacR-Dq complex crystals was required to obtain any crystals. These seeds generated multiple crystals. Crystal growth was limited, so sequential seeding was required to obtain crystals large enough for data collection (100 μm^3). These crystals were a dark red-brown color. Thus ethidium was incorporated into the crystal. Furthermore, the chance that any dequalinium remained in the crystal was slim as crystals from as single seed crystal and that the crystal that was studied was sequentially placed in two to three separate drops with high ethidium concentrations.

Please refer to Table 2 for further crystallographic and model statistics.

E120Q-QacR-MG, E120Q-QacR-Dq, and E120Q-QacR-R6G complex structures

The 2.4 Å resolution data of the structure of the E120Q-QacR-MG complex reveals that the E120Q-QacR-MG complex structure is the same as both WT-QacR-MG and E90Q-QacR-MG (Figure 2A). Furthermore, the imidazoles previously identified in the E90Q-QacR-MG complex structure appear in this structure (Figure 2B). The

presence of imidazoles in these structures is not surprising, as imidazole is a lipophilic cation at the low pH of the crystal conditions.

The 3.2 Å resolution E120Q-QacR-R6G and 2.9 Å resolution E120Q-QacR-Dq complex structures are also the same as their WT-QacR structures (Figure 3A&4). The E120Q-QacR-R6G complex structure has some extra electron density in the binding pocket that may be disordered solvent molecules (Figure 3B). Yet adding various known solutes in the cryosolvent (e.g. chloride and glycerol) did not increase the fit of the model, and no other orientations of R6G in the constraints of the pocket could account for this electron density. The density is likely to correspond to a few waters, as waters from the WT-QacR-R6G structure correspond to part of this electron density.

E120Q-QacR-Et complex structure

Similar to the E90Q-QacR-R6G complex crystal structure, the 3.3 Å resolution E120Q-QacR-Et complex structure has some unidentified density in the pocket. As a whole, the composite omit map indicates weak electron density in the binding pocket. There is both composite omit map and F_o-F_c map electron density that corresponds to the position of ethidium observed in WT-QacR-Et complex structure (Figure 4). However, including ethidium in the model in this position does not fit the data well, as the R_{free} increases by 0.1%, although R_{work} improves by 0.4%.

Another portion of the F_o-F_c map electron density in the pocket is closer to W61. The electron density in this position in the pocket suggests that there may be a molecule in a different portion of the binding pocket than the position of ethidium in the WT structure. It is possible that this electron density may correspond with a Dq remaining in the pocket, because E120Q-QacR-Et complex crystals were grown by seeding with

E120Q-QacR-Dq complex crystals. However, the extra electron density does not correspond with the aminomethylquinolinium head group of Dq, as Y103 is in a position that would exclude Dq from the binding pocket. Of the many positions tested, only one position for the ethidium improved the model fit (Figure 5). The model fitting statistics, R_{free} and R_{work}, were lowered by 0.04% and 0.68%, respectively.

The newly identified position for ethidium is pushed towards W61 by 6.3 Å and twisted to a position orthogonal to the wild-type position, while still overlapping significantly with the previous binding site (Figure 6A). The position of this ethidium is similar to what was seen in E90Q-QacR-Et (Figure 6B). The molecule of ethidium in each structure lies in the same plane, and about 1/3 of each molecule overlaps. However, the ethidium from the E120Q-QacR-Et complex structure is rotated about 120° in comparison to E90Q-QacR-Et. The position of Y103 would prevent the ethidium from taking the exact ethidium position from the E90Q-QacR-Et complex structure. Tyrosine 103 is in the orientation of the WT-QacR-Et complex structure rather than the E90Q-QacR-Et.

The interactions for this position have characteristics that are consistent with van der Waals or π-π interactions. The β-carbon from S86 is 3.4 Å from the terminal carbon of the ethyl group on ethidium and the γ-carbon from T89 is 3.6Å from the phenyl ring. The interactions, however, of the phenyl ring with W61 and Y93 at 3.5 Å and 3.7 Å respectively could be π-π interactions. The rings are almost perpendicular to each other; however, the edges are pointing towards each other, which points the positively charged portions of each ring system towards each other. Similarly Y123 at distance of 4 Å is interacting face-on-face with the ring system of the ethidium which is not the usual

conformation of a π-π interaction. Though this geometry may be optimal due to the delocalized positive charge in the ethidium. There may be a charge-dipole interaction between the δ-oxygen of N157 and one of the nitrogens on the end of the ethidium where the positive charge should be gathered. The least favorable interaction is with the nitrogen on the other side of the ethidium. Although this is the other center of positive charge in the molecule, the nitrogen is only 2.7 Å from the δ-carbon of I99.

Discussion

Role of E120Q for Ligand Affinity

The affinity of E120Q-QacR for ethidium is not much different then the affinity of QacR-WT for ethidium. Yet, this substituted protein increases the affinity for ethidium slightly. The two-fold change of affinity is equal to a ΔΔG of -0.4 kcal/mol. Considering that a hydrogen bond is approximately 5 kcal/mol, this affinity change might not be reflected structurally (*82*). However, π-cation and π-π interactions have energies in this range, so extra ligand-protein interactions with either of these types of interactions may account for the change, as there are a number of aromatic residues that contact the ligand in the pocket (*83*). Although the affinity change is consistent with the gain of a π-cation or π-π interaction, the crystal structure does not clearly indicate how a gain of affinity would be achieved. The geometry for the π-π interactions are not ideal, and in general, it is not clear that the ethidium in position 2 has better interactions with QacR than ethidium in position 1.

The negative charge of E120, like E90, does not play a direct role in the affinity of its cationic ligands. Furthermore, the E120Q substitution diminishes ethidium binding,

as the E120Q-QacR binding pocket has a slightly higher affinity for ethidium than does the WT-QacR binding pocket. The tyrosines provide a flexible binding surface that can interact with ligands through π-π, cation-π, aryl-sulfur, hydrogen bonding, and van der Waals interactions (*83*). The π-electrons of these aromatic residues make the faces of aromatic molecules partially negative and therefore contacts with aromatic residues may be all that is necessary to complement the positive charge of the ligand. There are also a number of hydrophobic and a few polar interactions available in the pocket. The interaction with E120 may be more important for molecules that do not take as much advantage of the other ligand binding interactions in the pocket.

The glutamate is also a part of the larger negative charge of the pocket. Other than the other glutamates, the main-chain carbonyls, and the π-electrons of aromatic residues provide a highly electronegative pocket. The electronegativity of the pocket may be all that is required for binding of the ligand. It is unfavorable to bury an uncomplemented positive charge in the dielectric of a protein. The global negative charge might remove this barrier and allow the interactions with the pocket to occur without specific interactions between negatively charged residues and the ligand. Further studies on the global charge of the binding pocket will ultimately determine if the negative charge of E120 has a role in ligand binding in the pocket, as well as the nature of such a role.

Multiple Binding Positions with Similar Affinity

The E120Q-QacR-Et complex structure has a novel binding position for ethidium. The E90Q-QacR-Et complex structure also has a novel position of ethidium binding. These positions overlap significantly but the difference in the position of Y103 in these

structures appears to be responsible for the 120° rotation between these ethidiums. The orientation of Y103 is the same in WT-QacR-Et and E120Q-QacR-Et complex structures. Thus, there is no change in the orientation of Y103 or any other tyrosine required for ethidium to take position 2 in this structure. Structurally, the ethidium in position 2 does not appear to interact with QacR any better than ethidium position 1. In fact, one of the nitrogens of the ethidium—a center of positive charge—is in a hydrophobic region of the binding pocket

Similar to the second position of E90Q it is difficult to explain the appearance of the new site for ethidium binding. If there were a change in the shape of the binding pocket, one could argue that somehow the substitution caused a change in the orientation of the residues in the pocket to form a higher-affinity binding pocket or to destabilize the old binding pocket. This new pocket could have a different affinity than the old pocket, and ethidium would shift to this new position. However, since the residues in E120Q-QacR-Et are not affected by the substitution, the binding site of position 2 has to be present with and without E120. If this is the case, ethidium should also occupy this position in the WT-QacR-Et structure. However, ethidium was only observed in one position in the wild-type structure.

It is possible that the electron density present in position 1 does represent an ethidium at this position, but the data is not strong enough to place the ligand in this position. The weak F_o-F_c map electron density and composite omit map electron density at the position 1 suggests there may be another ethidium at this site but at much lower occupancy than in position 2. The data here may be too weak to be fit. It is possible that like the two positions for R6G in the E90Q-QacR-R6G complex structure there are two

mutually exclusive ethidium binding positions in this pocket. At this resolution and lower occupancies it may not be possible to fit an ethidium in this density at all.

Influence of the conformation of the protein in the structure is not likely to be responsible for the shift in the location of ethidium binding. The E120Q-QacR-Et structure was solved in the $P6_2$ space group, compared to the $P4_22_12$ space group of the WT-QacR-Dq crystals. There is a conformational difference between the structures of the proteins in these two types of crystals. The different conformations of the protein could allow for different shapes of the ligand-binding pocket, and thus affect ligand binding. This conformational difference is not significant in the binding pocket: alignment of the regions of the WT-QacR-Et and E120Q-QacR-Et complex structures involved in ligand binding has an RMSD of 0.7 Å. The accessible volumes of the binding pockets of E90Q-QacR-Et and Wt-QacR-Et, as measured by the program VOIDOO, indicate a difference of only about 10 $Å^3$. Furthermore, using an overlay of the WT-QacR-Et and E120Q-QacR-Et structures, no structural changes that would favor one position of ethidium or the other could be found. Moreover, the E90Q-QacR-R6G structure also has two overlapping ligand positions in the binding pocket, compared with one ligand in the WT-QacR-R6G. In this case, the affinity does not change at all, and the space group of the E90Q-QacR-R6G crystal is the same as the WT-QacR-R6G complex, $P4_22_12$.

Summary

Like E90, E120 influences the positions of ligands in the binding pocket but it is not critical for their affinity. There is no evident relationship between the negative charge of the glutamate residue and the positive charges of the ligands with which it

interacts. Further studies on other glutamates in the binding pocket are required to see if these residues are also not critical for ligand binding. The individual charges of the glutamates may not be required as the binding pocket as a whole is highly electronegative and π-cation interactions with the aromatic residues may substitute for a formal negative charge of a glutamate.

The most significant finding provided by this research is that the QacR has multiple binding positions of the similar affinity for a ligand. This further illustrates how QacR is an ideal multidrug binding protein. Not only can QacR bind numerous drugs with high affinity, it can also bind the same drug in different ways with the same affinity. The degenerate ligand binding further illustrates how promiscuous the QacR ligand-binding pocket is.

Figure 1. ITC binding data for E120Q-QacR and Dq. The top panel is the raw multiple injection ITC binding data and the lower panel is the binding isotherm determined from that data. The average values for the experiments are included.

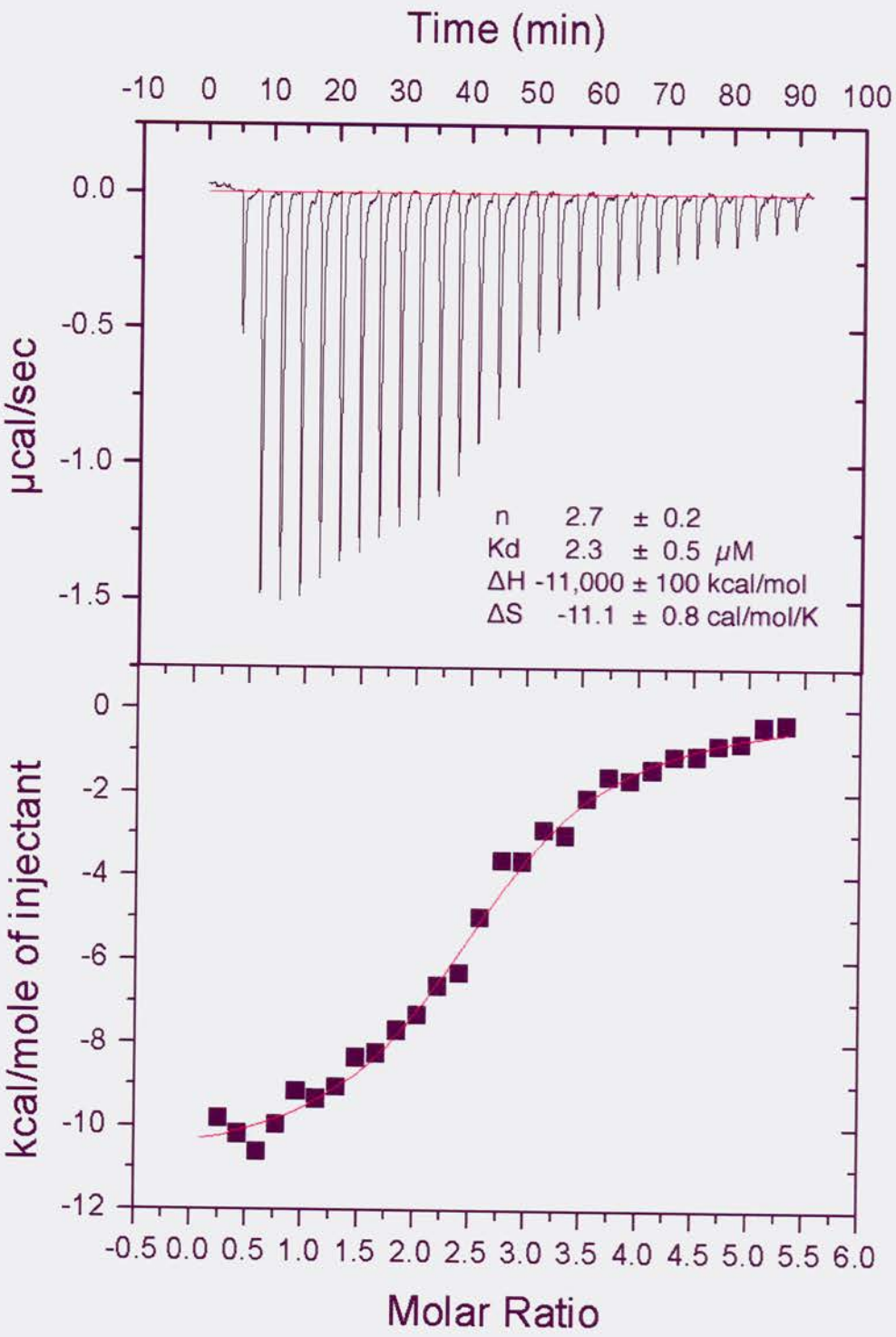
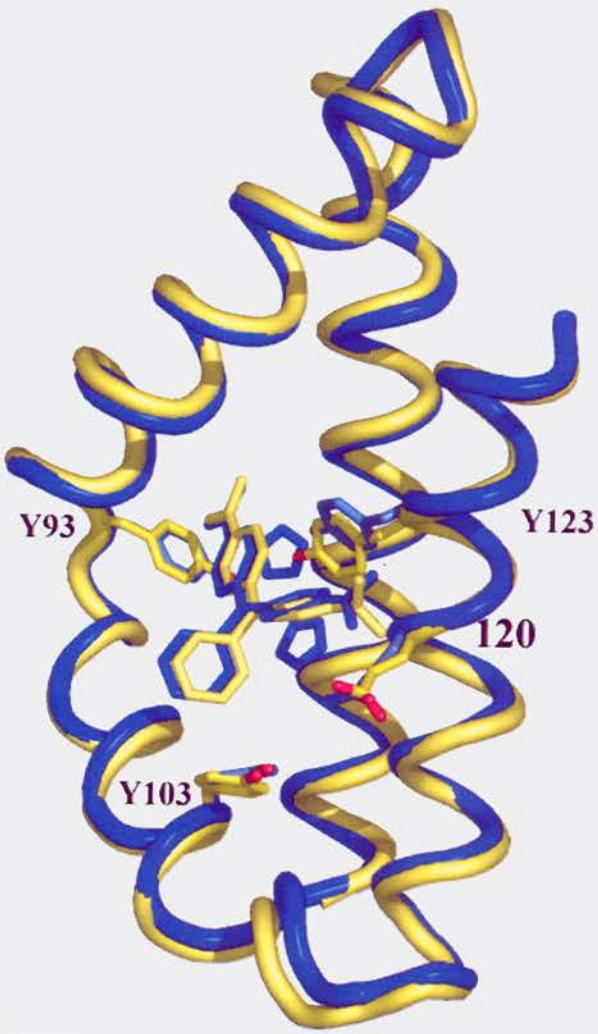
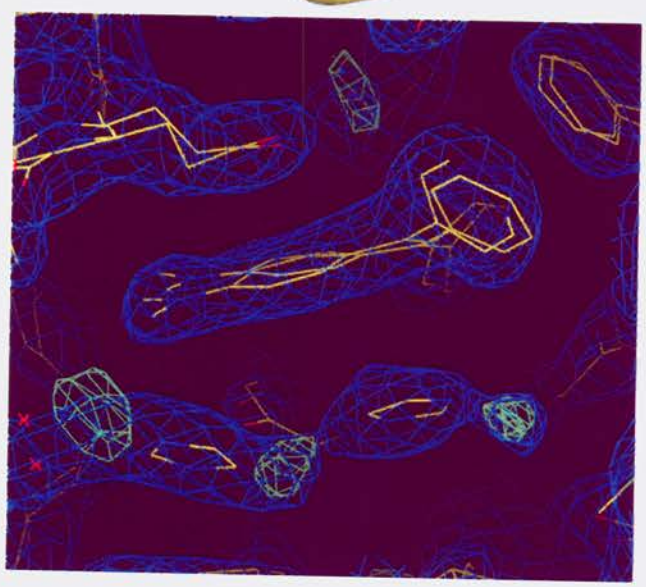

Time (min)
-10 0 10 20 30 40 50 60 70 80 90 100
0.0
-0.5
-1.0
-1.5
µcal/sec
n 2.7 ± 0.2
Kd 2.3 ± 0.5 µM
ΔH -11,000 ± 100 kcal/mol
ΔS -11.1 ± 0.8 cal/mol/K
0
-2
-4
-6
-8
-10
-12
kcal/mole of injectant
-0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Molar Ratio

Figure 2. E120Q-QacR-MG complex and WT-QacR-MG complex structures. A. The E90Q-QacR-MG complex (blue) and WT-QacR-MG complex (yellow) structures are identical except for the imidazoles that appear in the E90Q-QacR-MG complex structure. B. Imidazole and water in $2F_o$-F_c (blue) composite-omit-map (contoured at 1 σ) and the remaining F_o-F_c map (green) (contoured at 3 σ) electron density.

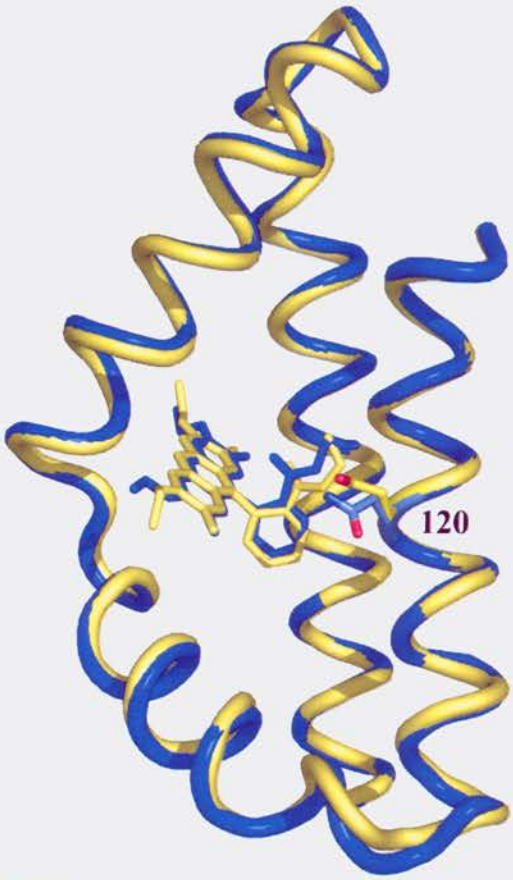


A

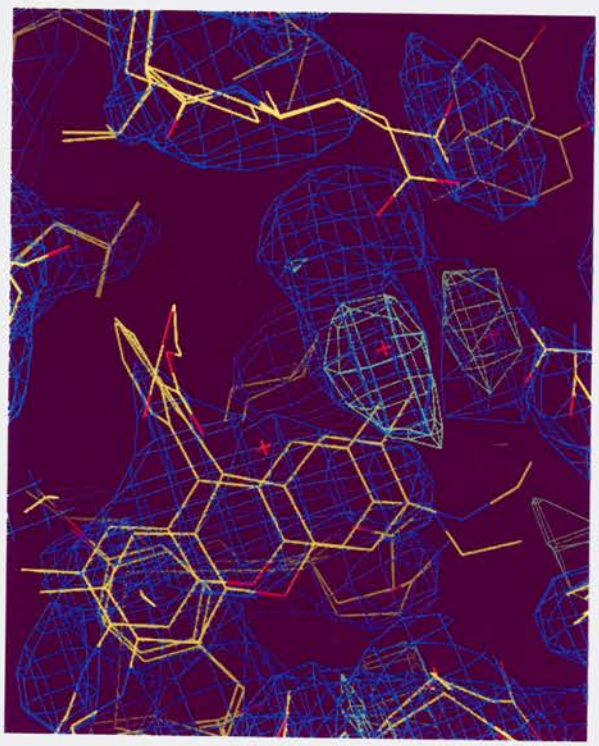


B

Figure 3. Comparison of R6G binding to the E120Q-QacR-R6G and WT-QacR-R6G complex structures. A. Overlay of E120Q-QacR-R6G (blue) and WT-QacR-R6G (yellow) complex structures. B. Electron density maps of E120Q-QacR-R6G binding pocket. Areas of F_o-F_c map (green)(contoured at 3 σ) and $2F_o$-F_c composite-omit-map (contoured at 1 σ) (blue) electron density probably correspond to water molecules.



A



B

Figure 4. Overlay of the WT-QacR-Dq complex and E90Q-QacR-Dq complex structures. WT-QacR-Dq complex is in yellow, and the E90Q-QacR-Dq complex is in blue.

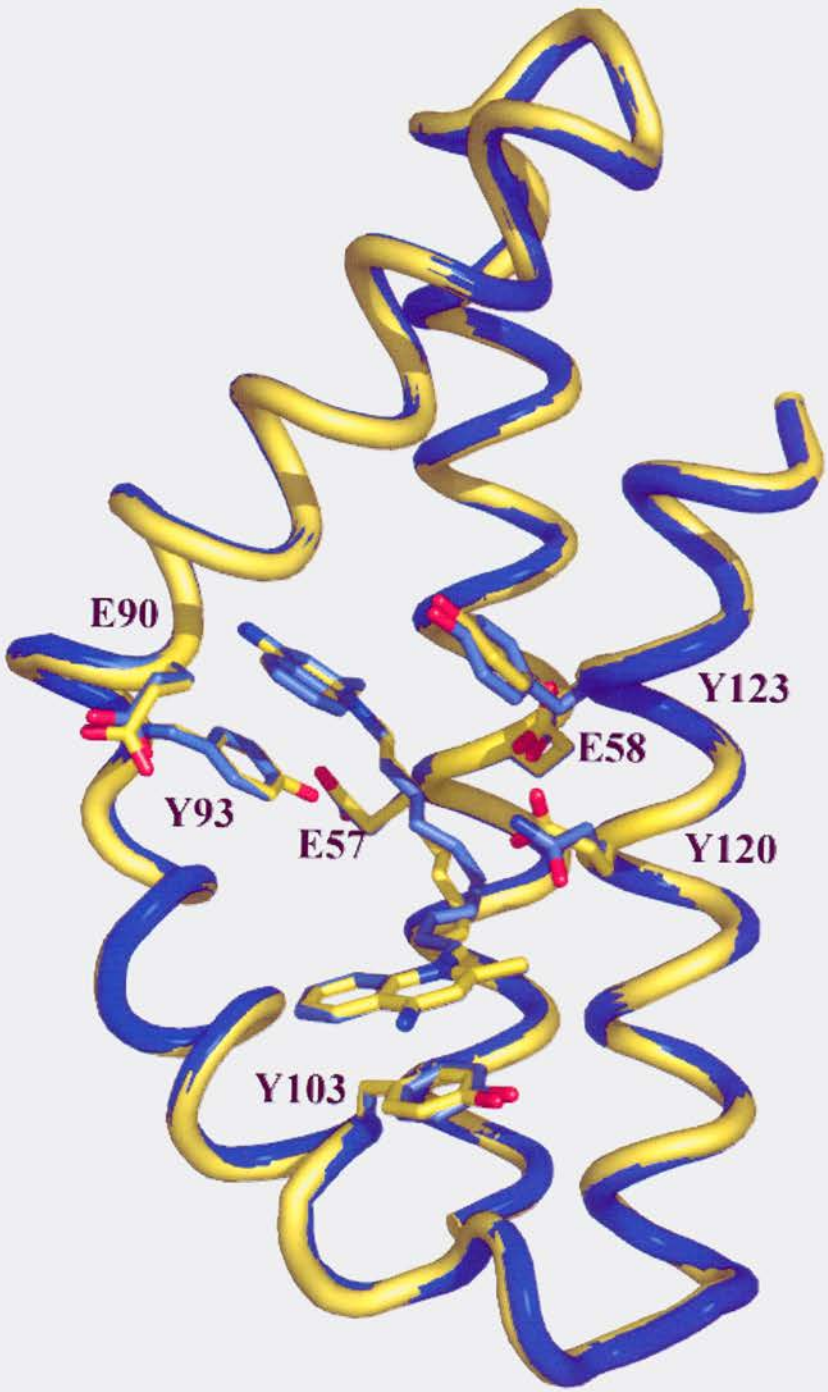

E90
Y123
E58
Y93
E57
Y120
Y103

Figure 5. Electron density in the E120Q-QacR-Et complex binding pocket. The electron density was contoured at 2.5 σ for the F_o-F_c map electron density and at 0.9 σ for the $2F_o$-F_c composite-omit-map electron density. The ethidium at the top of the figure is located at the position of ethidium from the WT-QacR-Et complex structure. The ethidium at the bottom is the position for ethidium revealed in the E120Q-QacR-Et complex structure.

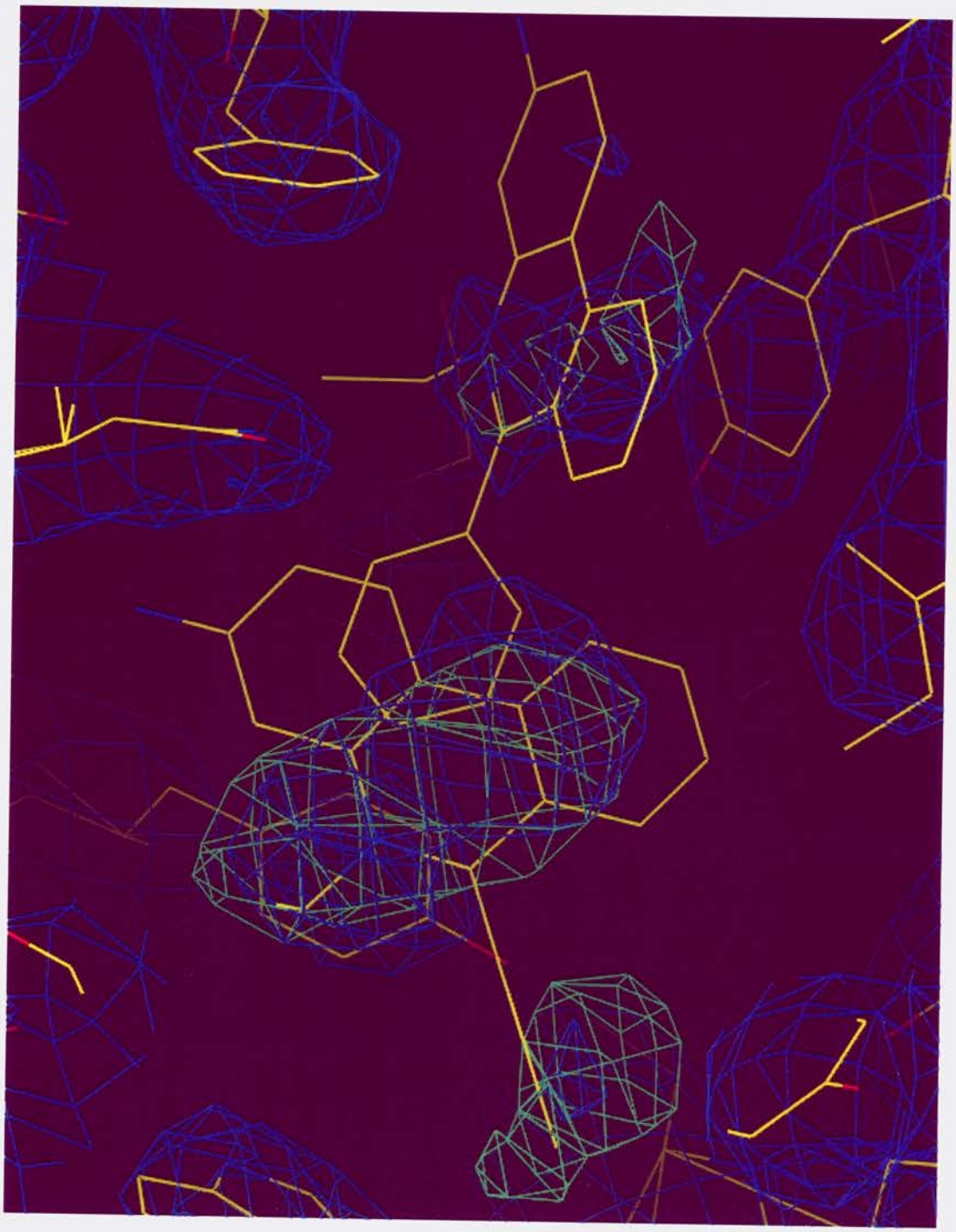
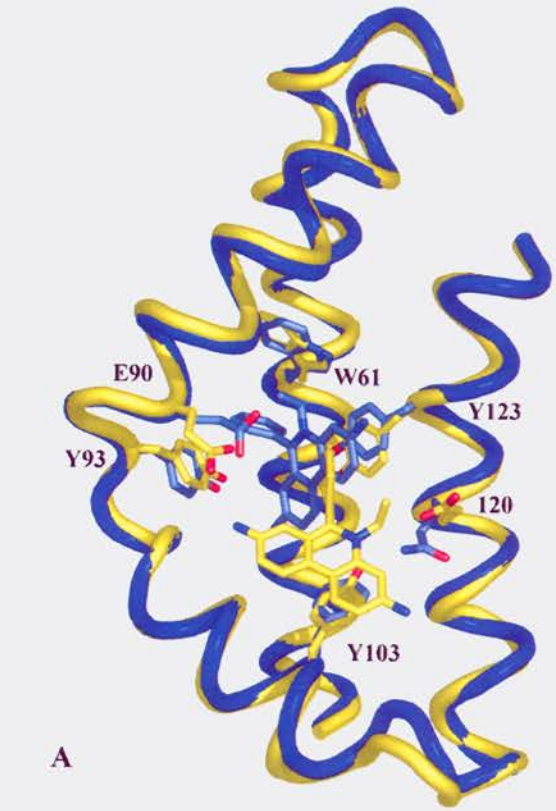


Figure 6. Comparison of the E120Q-QacR-Et binding position with ligands from other QacR-ligand complexes. A. Tyrosines 93, 103, and 123 of E120Q-QacR-Et complex (blue) are identical to those of WT-QacR-Et complex (yellow). The glutamates are disordered in the E120Q-QacR-Et structure. B. An overlay with E90Q-QacR-R6G (pink) illustrates how position 2 for ethidium in the E120Q-QacR-Et complex is different than the position in the E90Q-QacR-R6G complex structure.


E90
W61
Y123
Y93
120
Y103
A

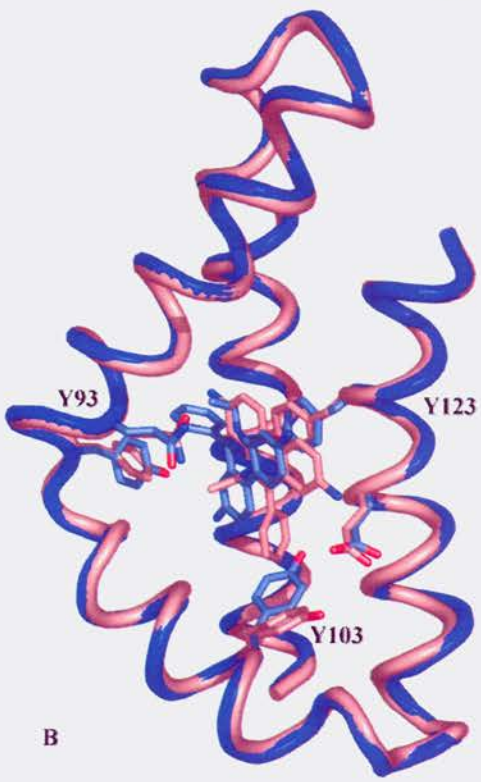

Y93
Y123
Y103
B

Table 1. E120Q-QacR dissociation constants.

	WT	E120Q
Et	2.4 ± 0.1 μM	1.18 ± 0.04 μM
MG	1.2 ± 0.1 μM	1.4 ± 0.1 μM
R6G	0.39 ± 0.02 μM	0.58 ± 0.04 μM
Dq	1.0 ± 0.5 μM	2.3 ± 0.5 μM

Table 2. X-ray diffraction and model statistics for E120Q-QacR-ligand structures

	E120Q-QacR-MG (K67S)	E120Q-QacR-Dq	E120Q-QacR-Et (K67S)	E120Q-QacR-R6G (K67S)
Temperature (K)	100	100	100	100
Space Group	$P6_2$	$P4_22_12$	$P6_2$	$P4_22_12$
Cell Constants (Å)	104.2	171.7	104.8	173.1
	104.2	171.7	104.8	173.1
	98.0	94.4	98.4	95.9
Resolution Range (Å) Low	52.1	48.9	46.3	49.4
High	2.4	2.9	3.3	3.2
CV-Luzzati Coord. Error	0.4	0.5	0.6	0.6
Overall Rsym (%)	8.5	7.1	8.1	5.9
Completeness	100	100	98.9	97.4
Overall I/sig(I)	6.1	5.9	5	6.7
Total Reflections	166250	237951	50185	75173
Unique reflections	23093	31749	9209	24147
Multiplicity	7.4	7.5	5.5	3.3
High Res. Shell Res.(Å)	2.6-2.4	3.0-2.9	3.4-3.3	3.3-3.2
Rsym(%)	43.5	52.7	47.5	42.2
I/Sig(I)	1.7	1.5	1.4	1.5
R_{work} (%)	21.8	22.8	23.1	23.7
R_{free} (%)	27.6	28.6	29.7	29.1
RMSD Bond Angles	1.1	1.4	1.2	1.3
Bond Lengths	0.008	0.010	0.009	0.010
Main Chain B Values	4.0	4.5	4.2	3.9
Solvent Water	76	31	10	11
Sulfate	11	16	5	15
Ramachandran Analysis				
Most Favored (%)	91.1	85.7	78.2	82.7
Additional Allowed (%)	6.9	12.4	19.5	16.3
Generously Allowed (%)	0.9	1.1	1.1	0.3
Disallowed (%)	1.1	0.7	1.1	0.7

Table 3. Contact Distances for Ethidium in E120Q-QacR-Et.

Ligand Atom	Partner Atom	E120Q (Å)
C1	N157 OD1	3.7
C2	N157 OD1	3.1
C3	N157 OD1	3.7
C4	Y123 CG	4.2
N5	Y123 CE2	4.0
C6	Y123 OH	3.7
	Y123 CZ	3.9
	Y123 CE2	4.0
C7	Y123 OH	4.1
C8	I99 CD1	3.4
	I99 CG1	4.0
	I99 CG2	4.2
C17	T89 CG2	3.7
C18	Y93 CE2	3.5
C19	W61 CB	3.3
C20	W61 CB	4.1
	W61 CG	4.1
C22	E90 OE2	3.0
	S86 O	3.8
	S86 CB	3.4
	N157 OD1	3.9
N23	N157 CB	3.8
	A153 O	4.2
N24	I99 CD1	2.8
	I99 CG1	3.2

Chapter 6: X-ray Crystallography Suggests QacR Has Multiple Binding Positions for Diamidine Compounds DB75 and DB 359

Introduction

QacR binds ligands in a variety of binding modes. A large binding pocket composed of two smaller pockets hosts a variety of differently shaped ligands in a variety of orientations. Ethidium and R6G bind differently, each to one of the smaller pockets(*1*). However, Dq and MG span both pockets but do so in distinct manners. Ethidium and MG interact with both E90 and E120. Dq interacts with E120 as well as E57 and 58, and R6G interacts only with E90. The ligands berberine and proflavin bind in the R6G pocket though they do not interact with E90 as R6G does. Berberine and proflavin interact with E57 and E58 rather than E90 (*8*).

The WT-QacR-pentamidine complex structure indicates a completely novel manner of QacR-ligand interaction. One end of the pentamidine leaves through a small passage proximal to W61 to the outside of QacR (Figure 1). At the protein's external surface, E63 contacts the positive charge of the pentamidine (*7*). The other end of the pentamidine inside the protein is complemented by the hydroxyl of tyrosine 127 rather than a glutamate. Interestingly, hexamidine, which is almost identical to pentamidine except for its slightly shorter length, binds to the structure in a manner similar to dequalinium, as hexamidine interacts with glutamates 57, 58, and 120 (Figure 1)(*7*).

To study whether the binding mode of pentamidine would be unique, we acquired two other diamidine type experimental drugs, DB75 and DB359 (courtesy of Dr. David Boykin, Department of Chemistry, Georgia State University). These compounds are

slightly smaller than pentamidine and are rigid, as, rather than an acyl chain, they include a furan ring conjugated between the two benzamidine head groups (Figure 2). Similar to pentamidine, DB75 is a potent DNA binding drug that exhibits topoisomerase II inhibition and indicates anti-giardial activity (*88; 88*). The different shape of DB359 does not allow for potent DNA binding (*89; 90*). I hypothesized that DB75 and DB359 would interact differently with QacR, and also that DB75 may bind similarly to pentamidine since it is similar to pentamidine. This compound however is shorter, bulkier, and more rigid than pentamidine and thus might bind in a completely novel manner.

Binding and structural studies of DB75 and DB359 with QacR indicate that these molecules bind in multiple ways to the binding pocket. The affinities of DB359 and DB75 by ITC are 15 and 18 μM respectively. The 2.8 Å X-ray crystal structures of the WT-QacR-DB75 and -DB359 complexes indicate that the protein is in the drug-bound state yet the ligand densities are completely disordered. This disorder is likely to be the result of competition between many similar binding positions in the larger binding pocket of the protein.

Results

DB75 & DB359 Binding

The affinity of QacR for these ligands was lower than for most QacR ligands. The affinities of most of the ligands as measured with fluorescence quenching are 2-3 μM (*6*). In comparison, the DB75 and DB359 have affinities of 18 ± 2 and 15 ± 0.2μM, respectively. This is similar to the binding affinity of other diamidines, as QacR's affinities for amicarbalide and phenamidine are 16 μM and 9 μM, respectively (*6*).

WT-QacR-DB75 and DB359 complex structures

Both the 2.8 Å WT-QacR-DB75 and the 2.8 Å WT-QacR-DB359 complex structures adopt the ligand bound conformation of QacR (Figure 3). Tyrosines 92 and 93 of the A subunit are no longer in the binding pocket. The DNA binding domains are moved away from each other, as they are in the induced form of the other QacR structures. The binding pocket of the B subunit remains empty of ligands and has Y92 and Y93 buried in the pocket.

Neither the WT-QacR-DB75 complex structure nor the WT-QacR-DB359 complex structure has any density that resembles or fits the drugs. WT-QacR-DB359 complex structure has a few continuous patches of electron density that looked promising for the ligand (Figure 4A). However, addition of a model of DB359 into this electron density did not fit the data well, as was reflected in the R-factor. Thus, the model for the WT-QacR-DB359 complex structure does not contain DB359. Similarly, the WT-QacR-DB75 complex structure did not have enough electron density to attempt addition of the drug to the model (Figure 4B). Thus, no model for DB75 was fitted into the WT-QacR-DB75 complex structure either.

Both the 2.8 Å WT-QacR-DB75 and the 2.8 Å WT-QacR-DB359 complex structures adopt the ligand bound conformation of QacR. Tyrosines 92 and 93 of the A subunit are no longer in the binding pocket (Figure 5A). The DNA binding domains are moved away from each other, as they are in the induced form of the other QacR structures. The binding pocket of the B subunit remains empty of ligands and has Y92 and Y93 buried in the pocket (Figure 5B).

The observed residues in the WT-QacR-DB359 and -DB75 complexes are in the same orientations as in the WT-QacR-Dq structure. However, a few of the binding pocket residues of the WT-QacR-DB75 complex structure are disordered. Modeling into the electron density suggests that E90 may have two conformations. Furthermore, is significant that Y132 is so disordered that no electron density is visible for it, as even lower resolution QacR structures (up to 3.3 Å) reveal ordered density for Y123s.

Discussion

Ligand Disorder Suggests Numerous Binding Positions

The ligand disorder in the WT-QacR-DB359 and DB75 complex structures illustrate that multiple sites in the ligand-binding pocket have similar affinities for each drug. The ligands are in so many different positions in the pocket that the electron density is lacking in the WT-QacR-DB359 and DB75 complex structures and a model cannot be fit to the structure. Other structures indicate that two or three positions for a ligand in a multidrug-binding pocket is possible. The E90Q-QacR-R6G and E120Q-QacR-Et complex structures both indicate two positions of the ligand in the binding pocket. Furthermore, the human multidrug-binding protein PXR complexed with drug SR12813 has three individual orientations of the ligand in the structure (*87*). In these structures, the electron density can become significantly weakened in certain areas. If there were more than a few positions of drug in the large QacR binding pocket, the electron density for the one compound would be spread out through the large binding pocket and would not be visible in the WT-QacR-DB359 and DB75 complex structures.

The multiple orientations taken by the residues in the pocket of the WT-QacR-ligand complexes help support this view. The unusual disorder of Y123 in WT-QacR-DB75 suggests that the residues in the pocket may be in many alternate orientations, these orientations accommodating the various positions of the drug in the binding pocket. Furthermore, glutamate 90 in its two orientations may also allow for variations in DB75 positioning in the pocket. The previous structures of QacR and PXR with various ligands indicate that different drugs can be accommodated in a multidrug-binding pocket by reorientation of residues in the pocket (*1; 7; 8; 91*). Thus, in the WT-QacR-DB75 complex structure, the residues in the binding pocket are moving to various positions to accommodate just one drug. This suggests, as have the E90Q-QacR-R6G, E90Q-QacR-Et, and E120Q-QacR-Et complex structures, that multiple sites of similar affinity are available in the larger ligand-binding pocket, often through remodeling of the surface of the ligand-binding pocket.

It is doubtful that the ligands are not visible due to the absence of ligand from the protein. All the previous structures with this ligand-bound form of the protein have ligand bound in the binding pocket. In these same structures, an apo-dimer of the protein is also included in the asymmetric unit. In this conformation, the "surrogate" ligands Y93 and Y92 occupy the binding pocket, thus excluding the possibility of ligand binding. The DB75 and DB359 also have a drug bound and apo dimer in their asymmetric units. In this case, the drug bound form can still be identified by the absence of Y92 and Y93 from the ligand-binding pocket. However, there is no electron density for the drug in the pocket. Thus, the drug is bound but not visible.

The disorder of the ligand in the pocket suggests that many binding positions in the pocket for these ligands have similar affinities. This seems plausible, as the surface of the binding pocket is malleable, because of all of the pivoting tyrosine side chains, and the ligand affinity is low compared to other QacR ligands. The E120Q-QacR-Et and the E90Q-QacR-R6G complexes have affinities close to 1 and 0.5 μM, respectively. Yet, the ligands in these structures are disordered. They are visible, as the ligands occupy only two positions. It may be possible that the ligands in the structures presented here are not visible at all, as they occupy many more than two positions in the binding pocket.

It is not necessarily the lower affinity of the ligands that is causing the disorder of the ligands in the crystal structure. Pentamidine, which has a much lower affinity than either DB75 or DB359, has electron density that is easily fit with a model (*7*). Furthermore, PXR has a much higher affinity of 41 nM for SR12813 and the PXR-LBD-SR12813 complex structure has three differently oriented molecules in the structure (*87*). As the ligands are generally in such high concentration, affinity itself is not necessarily the deciding factor for disorder. If there is only one site for the ligand, the site will likely be occupied, as is the case for pentamidine. However, ligands with higher affinities for QacR but that have multiple binding sites of similar affinity will likely be disordered in the structure.

Summary

Similar to the E120Q-QacR-Et and the E90Q-QacR-Et and R6G complex structures solved previously, the WT-QacR-DB359 and DB75 complex structures illustrate the multitudes of binding surfaces of the QacR binding pocket. This must allow for the great variety of structures that make up the ligands for this protein.

Figure 1. Pentamidine and hexamidine binding to QacR. Pentamidine (light green) interacts with E63 and the hydroxyl from Y127. Hexamidine (purple) and dequalinium (pink) interact with E57, E58, and E120.

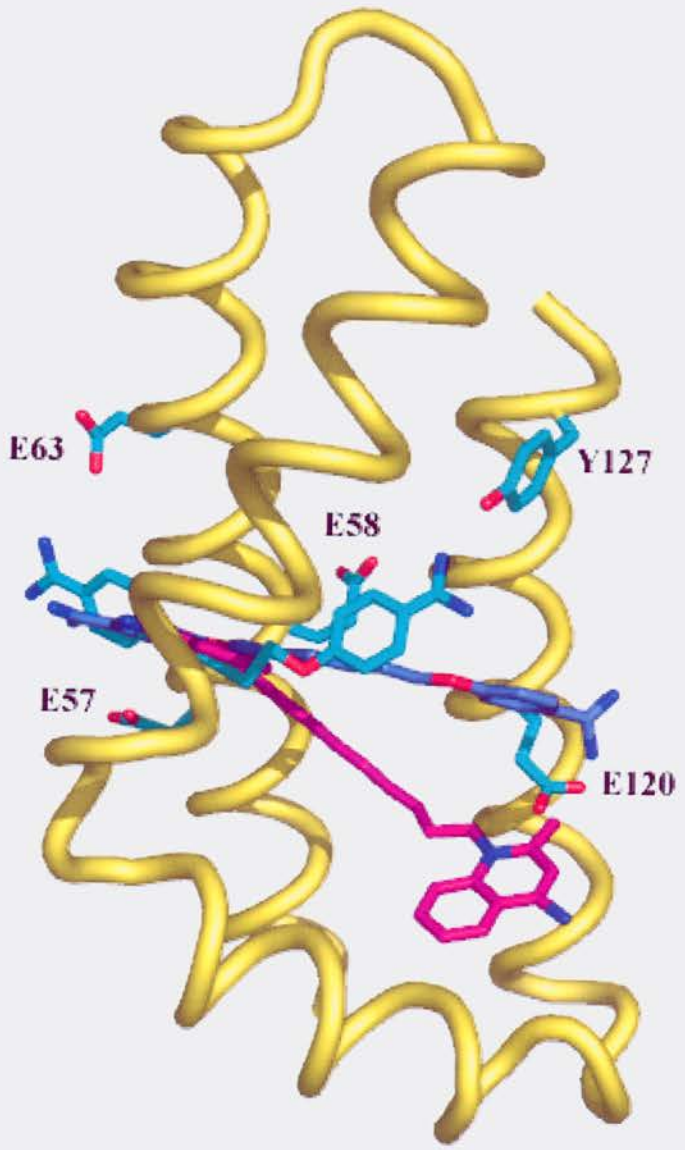
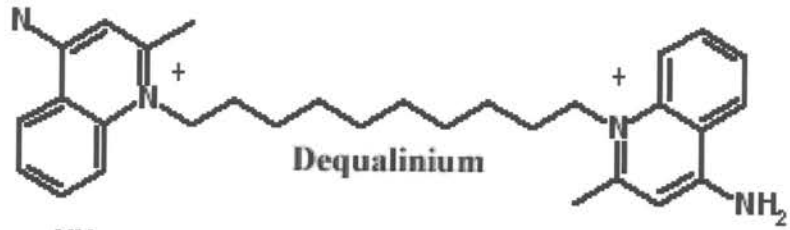
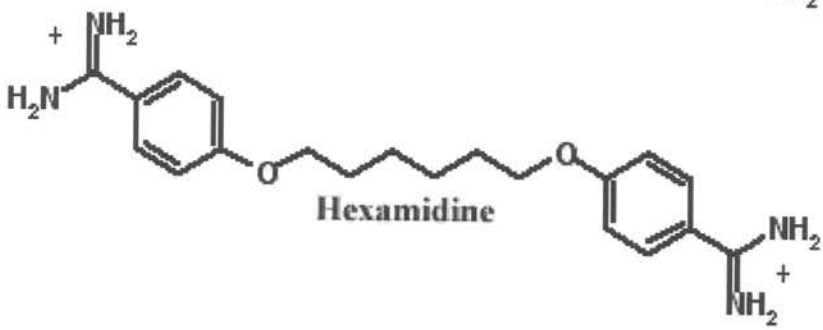

E63
Y127
E58
E57
E120

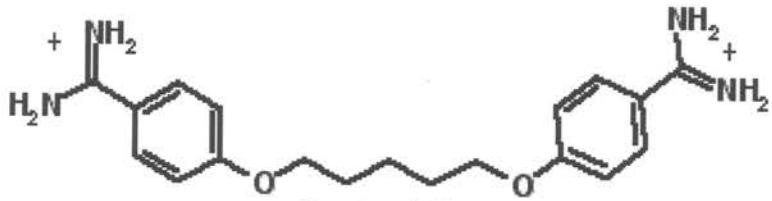
Figure 2. Comparison of diamidines.



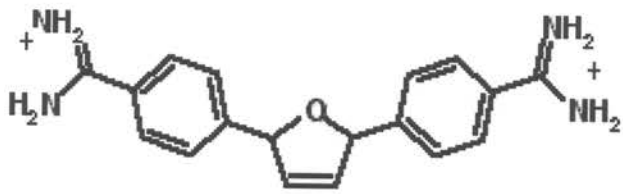
Dequalinium



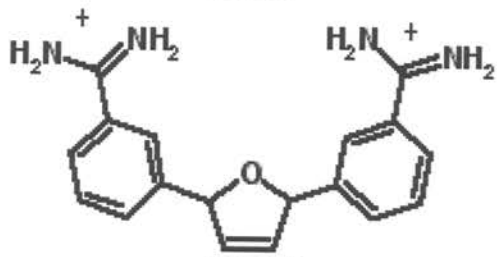
Hexamidine



Pentamidine



DB75



DB359

Figure 3. Overlay of the WT-QacR-DB75, WT-QacR-DB359, and WT-QacR-Dq complex structures. The WT-QacR-DB75, WT-QacR-DB359, and WT-QacR-Dq complex structures are blue, purple, and yellow respectively.

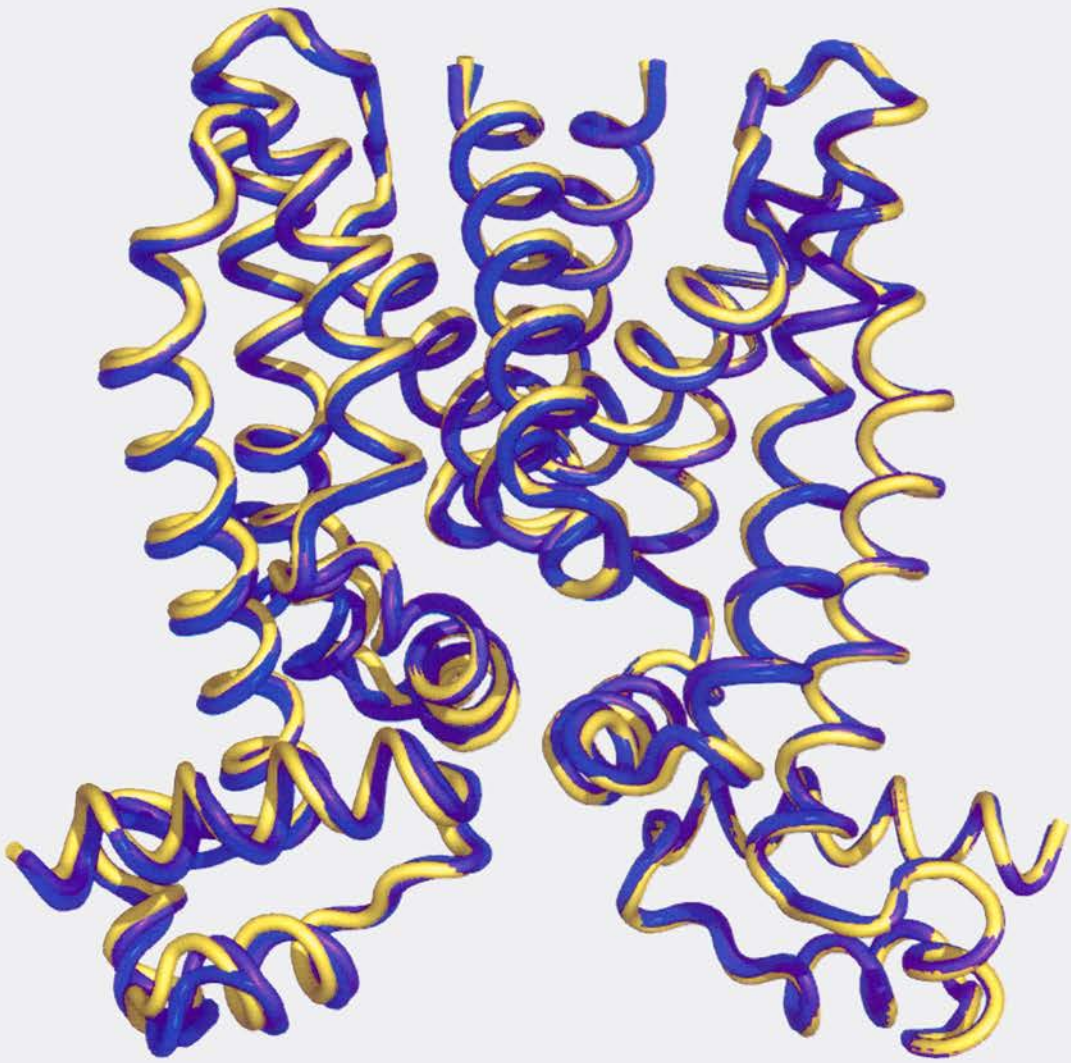


Figure 4. Electron density maps of DB359 and DB75. A. Electron density maps for DB359 with an overlay of WT-QacR-Dq indicate there is little continuous $2F_o$-F_c (blue) composite-omit map or F_o-F_c (green) map electron density into which to build. B. Electron density maps for DB75 indicate there is little continuous $2F_o$-F_c composite-omit map or F_o-F_c map electron density into which to build.

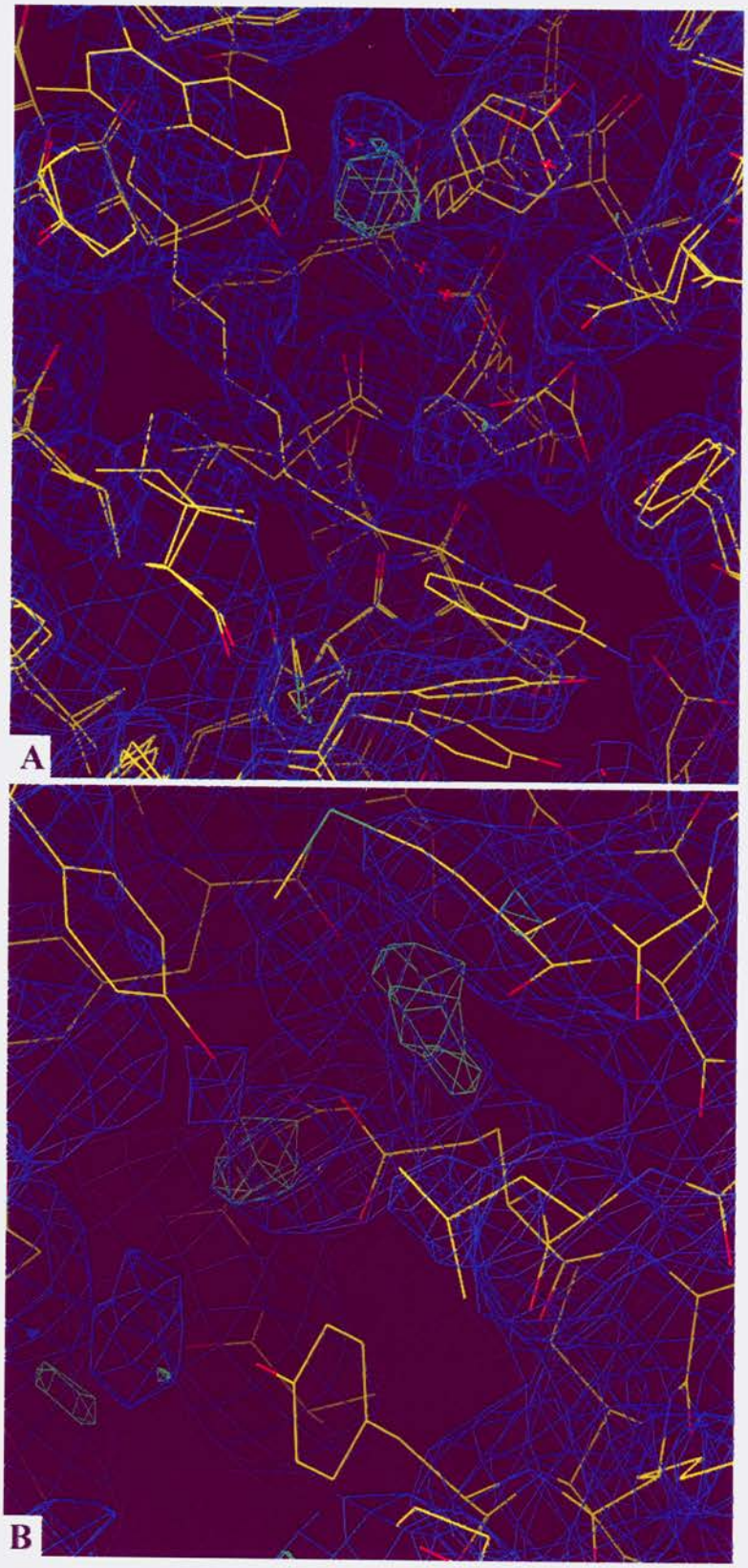

A
B

Figure 5. Overlay of the residues of the WT-QacR-DB359 and WT-QacR-DB7. The WT-QacR-DB359, WT-QacR-DB75, and WT-QacR-Dq complexes are lavender, cyan, and yellow respectively. A. The ligand binding pocket from the A subunits are identical. Tyrosine 123 from WT-QacR-DB75 (cyan) is disordered in the structure so its position is irrelevant. B. The residues from the B subunit are identical as well.

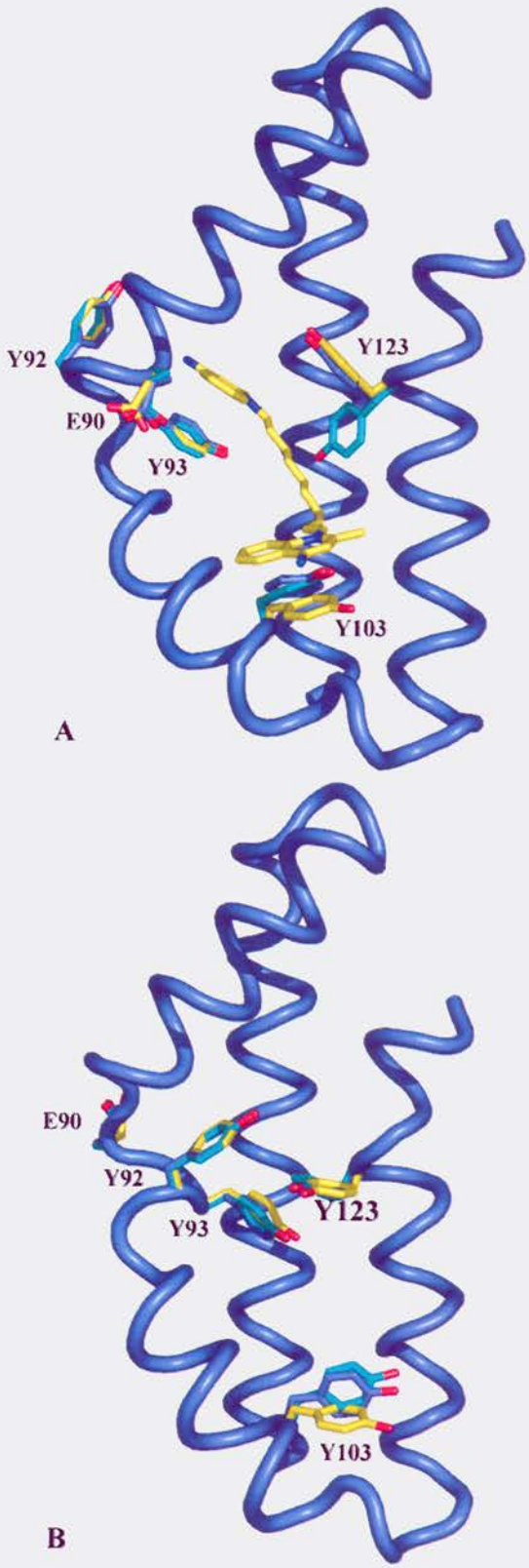

Y92
Y123
E90
Y93
Y103
A
E90
Y92
Y123
Y93
Y103
B

Table 1. X-ray diffraction and model statistics for WT-QacR-DB75/DB359 structures.

		WT-QacR-DB75	WT-QacR-DB359
Temperature (K)		100	100
Space Group		$P4_22_12$	$P4_22_12$
Cell Constants (Å)		172.0	171.4
		172.0	171.4
		94.6	94.5
Resolution Range (Å)	Low	81.7	81.7
	High	2.8	2.8
CV-Luzzati Coord. Error		0.52	0.50
Overall Rsym (%)		8.1	5.5
Completeness		98.5	96.9
Overall I/sig(I)		5.2	10
Total Reflections		116154	170971
Unique reflections		34500	10630
Multiplicity		3.37	5.09
High Res. Shell	Res.(Å)	3.0-2.8	3.0-2.8
	Rsym(%)	38	41
	I/Sig(I)	1.9	1.9
	R_{work} (%)	22.2	23.9
	R_{free} (%)	29.8	27.4
RMSD	Bond Angles	1.2	1.2
	Bond Lengths	0.008	0.008
	Main Chain B Values	1.6	1.3
Solvent	Water	6	24
	Sulfate	8	12
Ramachandran Analysis			
	Most Favored (%)	84.7	89.1
	Additional Allowed (%)	13.9	9.7
	Generously Allowed (%)	0.6	0.4
	Disallowed (%)	0.9	0.7

Chapter 7: Discussion

This dissertation further demonstrates how the QacR binding pocket binds a multitude of disparately shaped ligands. Not only does this research confirm that the binding pocket changes to accommodate various ligands, but it illustrates that the same ligand can interact with the protein in more than one way. The E90Q-QacR-R6G, E90Q-QacR-Et and E120Q-QacR-Et complex structures illustrate this well. The ligand binding is comparable to WT-QacR, yet a new position for ligand binding is observed in each of these structures. Furthermore, the ligands in the WT-QacR-DB75 and DB359 structures are likely to be taking many more than two positions in the binding pocket. Together these structures indicate that the binding pocket has many different surfaces that will act to bind ligands with similar affinities.

Like the WT-QacR-ligand complex structures, the residues in the structures presented in this dissertation reorient themselves to accommodate various drugs. However, some of these structures also show how reorientation of residues alters the surface of the binding pocket to create different binding sites for the same ligand. The orientation of Y103 alters between the WT-QacR-Et and E90Q-QacR-Et complex structures to accommodate the new position of ethidium in the structure. Glutamate 120 changes position to accommodate the alternate positions of R6G in the E90Q-QacR-R6G complex structure. Furthermore, the disorder of residue Y123 in the WT-QacR-DB75 complex structure suggests that this residue is adopting multiple orientations to create a variety of binding surfaces for the ligand. The plasticity of the binding pocket is reminiscent of the induced fit model of ligand binding in which the binding pocket is a surface that adjusts to the ligand. However, the E120Q-QacR-Et complex structure also

illustrates that no reorientation is required for a second binding site for a ligand, as ethidium is found in a different position and the tyrosines are unmoved compared to the WT-QacR-Et complex structure.

It is interesting that the new positions of taken by the ligands in these structures is not accompanied by a greater change in affinity of the protein for the ligand. One might expect that if the affinity does not change, then neither would the position of the ligand in the pocket. Furthermore, if the affinity for a second position is similar to the affinity for the position seen in the WT-QacR, one might expect that both positions for the ligand would be visible in the WT-QacR structure. However, the data clearly indicate that a substitution in QacR can affect the geometry of binding without affecting its affinity. This may be accounted for by the possibility that the glutamates may have roles in shaping the surface of the binding pocket. The glutamates may affect the orientations of tyrosines 123 and 103 and other residues that shape the binding pocket. These effects, however, would have to be felt over very long distances. Furthermore, the data of E120Q-QacR-Et complex structure does not support this model as it shows the ligand in two various positions without any alteration of the orientations of the residues in the binding pocket.

Clearly, removal of the formal-negative charge of glutamates 90 and 120 through substitution by glutamine has no significant deleterious effect on the affinity of ligands that contact these residues. I expected to see greater changes as charge-charge interactions seem an ideal candidate for interactions in this pocket. Given that the pocket is fairly hydrophobic and that the ligand itself is hydrophobic one would guess that the

dielectric of the pocket would be fairly low. This would allow for strong interactions between the glutamate and the ligand. But this does not appear to be the case.

There are a number of factors that might be partially responsible for this discrepancy. The dielectric constant is very difficult to estimate for ligand binding interactions. Furthermore, there are waters in the pocket. The glutamates in this pocket may not be fully deprotonated, especially in such a highly electronegative environment. As the charge of the ligand is spread over such a large area, the interaction with the glutamate is not likely to be as strong as if the charge was localized proximal to the glutamate. This delocalized positive charge may also interact with the other electronegative residues in the pocket thus decreasing the importance of the interaction with any particular glutamate.

Given that the surface of the binding pocket is extremely electronegative, it is unlikely that neutralizing the charge of one molecule would much affect the character of this pocket. It is possible that the overall charge of the pocket is more important than the local charges in the pocket. Thus, making individual favorable interactions with the ligands may not be the true role for the glutamates in the pocket. It is possible that rather than attracting the positive ligands to the pocket the negative charge of the pocket may be to allow the positive ligands to bind in a hydrophobic pocket. Perhaps the global negative charge of the pocket is enough to counter the positive charge of the ligand.

It should also be noted that E120Q-QacR was only tested with a very small subset of the disparate molecules it can bind. There may be other known ligands that do depend on the glutamates for ligand affinity. Furthermore, the natural substrates for the QacA/R system are unknown. Although the molecules we tested in these studies do not have a

dramatically reduced affinity due to these mutations, it may be because they are not interacting with the binding pocket as the true ligands would. The natural ligands may bind in a different orientation and have a different enough chemical composition that the glutamates may have an important role for their affinity to the binding pocket. Thus, there may be untested molecules that do have major changes in their affinity when the glutamates are substituted.

All the ligands tested here were aromatic; the positive charge of the ligand is likely to be distributed in the π system of the molecule. Thus all the interactions between the π-systems of the aromatics in the binding pocket and the ligand may be enough to counter the charge of the ligand. This cation-π interaction has often been observed in proteins but such a distributed π-system gives the protein more opportunity to counter the positive charge of the ligand with the constituents of its electronegative binding pocket. It would be interesting to try these substitution experiments with a QacR ligand that has a positive charge that is not connected to an aromatic system. A more localized positive charge might be more affected by changes in the local changes of the electrostatic nature of the pocket.

The exact nature of these ligands is not clear, nor is it clear how to predict them to interact with the π-systems. The ligands are π-systems themselves, as well as cations. Furthermore, the positive charge of the ligands is distributed across the π-system of the ligand. Interactions between the π-systems of aromatic residues (*π-π interactions*) in proteins are usually offset or perpendicular. In QacR structures, we observe both face-to-face and perpendicular interactions by the aromatic groups in the binding pocket. It is not

clear if these interactions are some kind of hybrid of π-π and cation-π interactions or if they are individual cases of one or the other.

The ligands may have an uneven distribution of the positive charge in its π-system. Parts of the π-system of the ligand may interact differently from others. Computer aided analysis of the charge distribution of these ligands and prediction of how they would interact with residues in the protein may be helpful for understanding exactly what contribution one would expect for different types of ligand-residue interactions. This would aid us in looking at the binding pocket to determine exactly what types of interactions we see.

The overall electronegative aspect of the binding pocket may have other purposes as well. Not only may it help bind ligands by either increasing the affinity of the ligand for the pocket or by countering the free positive charge of the ligand, but it may also screen out negatively charged molecules. If this were the case, it would seem that neutral lipophilic molecules would be decent ligands. However, the ligands are solely cationic. To simultaneously test if the overall negative charge screens out negative ligands and whether the negative charge attracts cations, one could take a monocationic QacR ligand and find or synthesize the same molecule but with a single negative charge and try binding them to the WT-QacR and the E78Q-E58Q-E90Q-E120Q-QacR. If the total negative charge of the glutamates is important for binding cations, then affinity for the positive ligand should decrease after the substitution of the glutamates. If the negative charge screens out the anions, then the affinity for the negatively charged version of the ligand should increase after the substitution of the glutamates. Both may be true. I have tried this experiment with derivatives of rhodamine 123, which is cationic. However, this

was unsuccessful as rhodamine 123, which of the molecules tested was the molecule most like R6G, was barely measurable. Thus, it was not surprising that the other ligands could not be measured.

I also attempted to measure the affinity of these rhodamine 123 derivatives as inhibitors of R6G binding. However, the high concentrations required and efficient fluorescence of the molecules hindered the experiments. Thus, the creation of differently charged ligands with the same shape that are not fluorescent may be helpful for studying the roles of the glutamates. The affinities of these differently charged molecules may either be measured by ITC or if the affinity is too low, the affinity of these molecules may easily studied with fluorescence polarization by studying the inhibition of R6G binding by these ligands.

The innate lipophilicity of these ligands poses a problem for ligand binding as well. These ligands bind to just about everything. This poses a problem with binding assays, as they might bind to the outside of the protein and the experimental apparati. Generally there are ways of measuring or counteracting these issues, but with the lower affinity interactions observed with QacR it may be helpful to be able to control for the ligand promiscuity. One possibility would be to create a mutant of QacR that does not bind any molecule in its binding site as a control for non-specific binding. This could be achieved by adding a bulky positively charged residue in the pocket such as an arginine. In this way, one could eliminate the possibility of ligand binding by adding a positive charge, filling the pocket with a hydrophobic chain, and sequestering the tyrosines through π-cation interactions. This QacR mutant may also be interesting from the point of view of binding mechanics as it should be constitutively induced. Finding a second

site mutation that would allow DNA binding could help identify other important residues for QacR induction.

The imidazole and other solute molecules in the pocket suggest possibilities for the pocket as well. Imidazole at the pH of the crystals would be a lipophilic cation and match the ligand profile for the protein. It is possible, however, that there are unidentified counter ions accompanying the ligand into the binding pocket. An example of this in a related structure occurs in the interaction of tetracycline and the TetR protein, as tetracycline binds to TetR as a complex with Mg^{+2}. At the resolution of these structures, if small counter ions enter the binding pocket of QacR with the ligands, they might not be observed or perhaps observed but not identifiable. At the synchrotron, one could test if other counter ions such as chloride are present by scanning for nuclear fluorescence.

One last aspect of these results to consider is how they might reflect on the mechanism ligand binding by QacA. Although we can use results and our understanding of the binding mechanism of QacR to create hypotheses concerning QacA ligand interactions, what we learn from QacR cannot exclude possible mechanisms for ligand binding by QacA. These two proteins are in entirely different dielectric environments so it is likely the role of charges in these proteins will be different.

In summary, this dissertation illustrates how promiscuous this ligand-binding pocket is. The properties of QacR make it an ideal model of how one would design a protein to interact with multiple differently shaped drugs and antiseptics. Exactly what roles the negative charges play in the pocket is still not clear. They may play an attractive and a repulsive role for ligand selection. Further studies of glutamates in the

pocket will illustrate if this is true and if the observations made in these studies remain true for all the other glutamates.

References

1. Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A., and Brennan, R. G. (2001) Structural mechanisms of QacR induction and multidrug recognition, *Science 294*, 2158-2163.
2. Levy, S. B., and Marshall, B. (2004) Antibacterial resistance worldwide: causes, challenges and responses, *Nat Med 10*, S122-9.
3. Weinstein, R. A. (2001) Controlling antimicrobial resistance in hospitals: infection control and use of antibiotics, *Emerg Infect Dis 7*, 188-192.
4. Hiramatsu, K. (1998) Vancomycin resistance in staphylococci, *Drug Resist Updat 1*, 135-150.
5. Noguchi, N., Suwa, J., Narui, K., Sasatsu, M., Ito, T., Hiramatsu, K., and Song, J. H. (2005) Susceptibilities to antiseptic agents and distribution of antiseptic-resistance genes qacA/B and smr of methicillin-resistant Staphylococcus aureus isolated in Asia during 1998 and 1999, *J Med Microbiol 54*, 557-565.
6. Grkovic, S., Hardie, K. M., Brown, M. H., and Skurray, R. A. (2003) Interactions of the QacR multidrug-binding protein with structurally diverse ligands: implications for the evolution of the binding pocket, *Biochemistry 42*, 15226-15236.
7. Murray, D. S., Schumacher, M. A., and Brennan, R. G. (2004) Crystal structures of QacR-diamidine complexes reveal additional multidrug-binding modes and a novel mechanism of drug charge neutralization, *J Biol Chem 279*, 14365-14371.
8. Schumacher, M. A., Miller, M. C., and Brennan, R. G. (2004) Structural mechanism of the simultaneous binding of two drugs to a multidrug-binding protein, *EMBO J 23*, 2923-2930.

9. Lim, D., and Strynadka, N. C. (2002) Structural basis for the beta lactam resistance of PBP2a from methicillin-resistant Staphylococcus aureus, *Nat Struct Biol 9*, 870-876.

10. Jordheim, L. P., Cros, E., Gouy, M. H., Galmarini, C. M., Peyrottes, S., Mackey, J., Perigaud, C., and Dumontet, C. (2004) Characterization of a gemcitabine-resistant murine leukemic cell line: reversion of in vitro resistance by a mononucleotide prodrug, *Clin Cancer Res 10*, 5614-5621.

11. Paulsen, I. T., Brown, M. H., Littlejohn, T. G., Mitchell, B. A., and Skurray, R. A. (1996) 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.

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

13. Neyfakh, A. A. (1992) The multidrug efflux transporter of Bacillus subtilis is a structural and functional homolog of the Staphylococcus NorA protein, *Antimicrob Agents Chemother 36*, 484-485.

14. 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.

15. Mayer, S., Boos, M., Beyer, A., Fluit, A. C., and Schmitz, F. J. (2001) Distribution of the antiseptic resistance genes qacA, qacB and qacC in 497 methicillin-resistant

and -susceptible European isolates of Staphylococcus aureus, *J Antimicrob Chemother 47(6)*, 896-897.

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

17. Mitchell, B. A., Paulsen, I. T., Brown, M. H., and Skurray, R. A. (1999) Bioenergetics of the staphylococcal multidrug export protein QacA. Identification of distinct binding sites for monovalent and divalent cations, *J Biol Chem 274*, 3541-3548.

18. Paulsen, I. T., Brown, M. H., and Skurray, R. A. (1998) Characterization of the earliest known Staphylococcus aureus plasmid encoding a multidrug efflux system, *J Bacteriol 180*, 3477-3479.

19. Mitchell, B. A., Brown, M. H., and Skurray, R. A. (1998) QacA multidrug efflux pump from Staphylococcus aureus: comparative analysis of resistance to diamidines, biguanidines, and guanylhydrazones, *Antimicrob Agents Chemother 42*, 475-477.

20. Grkovic, S., Brown, M. H., and Skurray, R. A. (2002) Regulation of bacterial drug export systems, *Microbiol Mol Biol Rev 66*, 671-701.

21. Jack, D. L., Yang, N. M., and Saier, M. H. J. (2001) The drug/metabolite transporter superfamily, *Eur J Biochem 268*, 3620-3639.

22. Rhee, S., Martin, R. G., Rosner, J. L., and Davies, D. R. (1998) A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator, *Proc Natl Acad Sci U S A 95*, 10413-10418.

23. Martin, R. G., and Rosner, J. L. (2001) The AraC transcriptional activators, *Curr Opin Microbiol 4*, 132-137.

24. Kwon, H. J., Bennik, M. H., Demple, B., and Ellenberger, T. (2000) Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA, *Nat Struct Biol 7*, 424-430.

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

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

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

28. Heldwein, E. E., and Brennan, R. G. (2001) Crystal structure of the transcription activator BmrR bound to DNA and a drug, *Nature 409*, 378-382.

29. Zheleznova, E. E., Markham, P. N., Neyfakh, A. A., and Brennan, R. G. (1999) Structural basis of multidrug recognition by BmrR, a transcription activator of a multidrug transporter, *Cell 96*, 353-362.

30. Newberry, K. J., Huffman, J. L., Miller, M. C., Vasquez-Laslop, N., Neyfakh, A. A., and Brennan, R. G. BmrR requires a buried glutamate for transcription regulation but not high affinity drug binding, *in press*

31. Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R., and Tobes, R. (2005) The TetR family of transcriptional repressors, *Microbiol Mol Biol Rev 69*, 326-356.

32. Natsume, R., Ohnishi, Y., Senda, T., and Horinouchi, S. (2004) Crystal structure of a gamma-butyrolactone autoregulator receptor protein in Streptomyces coelicolor A3(2), *J Mol Biol 336*, 409-419.

33. Frenois, F., Engohang-Ndong, J., Locht, C., Baulard, A. R., and Villeret, V. (2004) Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis, *Mol Cell 16*, 301-307.

34. Eckert, B., and Beck, C. F. (1989) Overproduction of Transposon Tn*10*-Encoded Tetracycline Resistanc Protein Results in Cell Death and Loss of Membrane Potential, *J. Bacteriol. 171*, 3557-3559.

35. Baron, U., and Bujard, H. (2000) Tet repressor-based system for regulated gene expression in eukaryotic cells: principles and advances, *Methods Enzymol 327*, 401-421.

36. Baron, U., Gossen, M., and Bujard, H. (1997) Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential, *Nucleic Acids Res 25*, 2723-2729.

37. Rossi, F. M., Guicherit, O. M., Spicher, A., Kringstein, A. M., Fatyol, K., Blakely, B. T., and Blau, H. M. (1998) Tetracycline-regulatable factors with distinct dimerization domains allow reversible growth inhibition by p16, *Nat Genet 20*, 389-393.

38. Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system, *Nat Struct Biol 7*, 215-219.

39. Grkovic, S., Brown, M. H., Hardie, K. M., Firth, N., and Skurray, R. A. (2003) Stable low-copy-number Staphylococcus aureus shuttle vectors, *Microbiology 149*, 785-794.

40. Grkovic, S., Brown, M. H., Schumacher, M. A., Brennan, R. G., and Skurray, R. A. (2001) The staphylococcal QacR multidrug regulator binds a correctly spaced operator as a pair of dimers, *J Bacteriol 183*, 7102-7109.

41. Dover, L. G., Corsino, P. E., Daniels, I. R., Cocklin, S. L., Tatituri, V., Besra, G. S., and Futterer, K. (2004) Crystal structure of the TetR/CamR family repressor Mycobacterium tuberculosis EthR implicated in ethionamide resistance, *J Mol Biol 340*, 1095-1105.

42. Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A., and Brennan, R. G. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR, *EMBO J 21*, 1210-1218.

43. Hoffmann, K. M., Williams, D., Shafer, W. M., and Brennan, R. G. (2005) Characterization of the multiple transferable resistance repressor, MtrR, from Neisseria gonorrhoeae, *J Bacteriol 187*, 5008-5012.

44. Engohang-Ndong, J., Baillat, D., Aumercier, M., Bellefontaine, F., Besra, G. S., Locht, C., and Baulard, A. R. (2004) EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator, *Mol Microbiol 51*, 175-188.

45. Merickel, A., Rosandich, P., Peter, D., and Edwards, R. H. (1995) Identification of residues involved in substrate recognition by a vesicular monoamine transporter, *J Biol Chem 270*, 25798-25804.

46. Yamaguchi, A., O'yauchi, R., Someya, Y., Akasaka, T., and Sawai, T. (1993) Second-site mutation of Ala-220 to Glu or Asp suppresses the mutation of Asp-285 to Asn in the transposon Tn10-encoded metal-tetracycline/H+ antiporter of Escherichia coli, *J Biol Chem 268*, 26990-26995.

47. Lemieux, M. J., Huang, Y., and Wang, D. N. (2004) Glycerol-3-phosphate transporter of Escherichia coli: structure, function and regulation, *Res Microbiol 155*, 623-629.

48. 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.

49. Paulsen, I. T., and Skurray, R. A. (1993) Topology, structure and evolution of two families of proteins involved in antibiotic and antiseptic resistance in eukaryotes and prokaryotes--an analysis, *Gene 124*, 1-11.

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

51. Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, *Nature 171*, 737-738.

52. Watson, J. D., and Crick, F. H. (1953) Genetical implications of the structure of deoxyribonucleic acid, *Nature 171*, 964-967.

53. Franklin, R. E., and Gosling, R. G. (1953) Molecular Configuration in Sodium Thymonucleate, *Nature 171*, 740-741.

54. Kendrew, J. C., Bodo, G., Dintis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis, *Nature 181*, 662-666.

55. Stout, G. H., and Jensen, L. H. (1989) X-Ray Structure Determination: A Practical Guide, 2nd Edition,

56. Drenth, J. (2002) Principles of Protein X-ray Crystallography (Springer Advanced Texts in Chemistry),

57. Yeates, T. O., and Fam, B. C. (1999) Protein crystals and their evil twins, *Structure Fold Des 7*, R25-9.

58. Yeates, T. O. (1997) Detecting and overcoming crystal twinning, *Methods Enzymol 276*, 344-358.

59. Ramakrishnan, V., and Biou, V. (1997) Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement, *Methods Enzymol 276*, 538-557.

60. Vagin, A. A., and Isupov, M. N. (2001) Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps, *Acta Crystallogr D Biol Crystallogr 57*, 1451-1456.

61. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) Rapid automated molecular replacement by evolutionary search, *Acta Crystallogr D Biol Crystallogr 55*, 484-491.

62. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations, *FEBS Lett 229*, 317-324.

63. Engh, R. A., and Huber, R. (1991) Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement, *Acta Crystallogr A 47*,

64. Silva, A. M., and Rossmann, M. G. (1985) *Acta Crystallogr B 41*,

65. Pannu, N. S., and Read, R. J. (1996) Improved Structure Refinement Through Maximum Likelihood, *Acta Crystallogr A 52*, 659-668.

66. Adams, P. D., Pannu, N. S., Read, R. J., and Brunger, A. T. (1997) Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement, *Proc Natl Acad Sci U S A 94*, 5018-5023.

67. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr D Biol Crystallogr 54*, 905-921.

68. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr D Biol Crystallogr 50*, 760-763.

69. 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.

70. Rayment, I. (1997) Reductive alkylation of lysine residues to alter crystallization properties of proteins, *Methods Enzymol 276*, 171-179.

71. Pflugrath, J. W. (1999) The finer things in X-ray diffraction data collection, *Acta Crystallogr D Biol Crystallogr 55*, 1718-1725.

72. McPhillips, T. M., McPhillips, S. E., Chiu, H. J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M., and Kuhn, P. (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines, *J Synchrotron Radiat 9*, 401-406.

73. GJ, K., and TA, J. (1998) Databases in Protein Crystallography, *Acta Crystallogr D 54*, 1119-1131.

74. TA, J., JY, Z., SW, C., and M, K. (1991) Improved methods for building protein models in electron density maps and the

location of errors in these models, *Acta Crystallogr A 47*, 110-119.

75. GJ, K., and TA, J. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures, *Acta Crystallogr D 50*, 178-185.

76. GJ, K., JY, Z., M, K., and TA, J. (2001) *International Tables for Crystallography F*, pp353-356, 366-367.

77. Lundblad, J. R., Laurance, M., and Goodman, R. H. (1996) Fluorescence polarization analysis of protein-DNA and protein-protein interactions, *Mol Endocrinol 10*, 607-612.

78. Leavitt, S., and Freire, E. (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry, *Curr Opin Struct Biol 11*, 560-566.

79. Holdgate, G. A. (2001) Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics, *Biotechniques 31*, 164-6, 168, 170 passim.

80. Jelesarov, I., and Bosshard, H. R. (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition, *J Mol Recognit 12*, 3-18.

81. Du, H., Fuh, R. A., Li, J., Corkan, A., and Lindsey, J. S. (1998) PhotochemCAD: A computer-aided design and research tool in photochemistry, *Photochemistry and Photobiology 68*, 141-142.

82. Holde, K. E. V. (1985) Physical Biochemistry,

83. Waters, M. L. (2004) Aromatic interactions in peptides: impact on structure and function, *Biopolymers 76*, 435-445.

84. Vincent, F., Ramoni, R., Spinelli, S., Grolli, S., Tegoni, M., and Cambillau, C. (2004) Crystal structures of bovine odorant-binding protein in complex with odorant molecules, *Eur J Biochem 271*, 3832-3842.

85. Bianchet, M. A., Bains, G., Pelosi, P., Pevsner, J., Snyder, S. H., Monaco, H. L., and Amzel, L. M. (1996) The three-dimensional structure of bovine odorant binding protein and its mechanism of odor recognition, *Nat Struct Biol 3*, 934-939.

86. Tegoni, M., Ramoni, R., Bignetti, E., Spinelli, S., and Cambillau, C. (1996) Domain swapping creates a third putative combining site in bovine odorant binding protein dimer, *Nat Struct Biol 3*, 863-867.

87. Watkins, R. E., Wisely, G. B., Moore, L. B., Collins, J. L., Lambert, M. H., Williams, S. P., Willson, T. M., Kliewer, S. A., and Redinbo, M. R. (2001) The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity, *Science 292*, 2329-2333.

88. Boykin, D. W., Kumar, A., Xiao, G., Wilson, W. D., Bender, B. C., McCurdy, D. R., Hall, J. E., and Tidwell, R. R. (1998) 2,5-Bis[4-(N-alkylamidino)phenil]furans as Anti-*Pneumocystic carnii* Agents, *J. Med. Chem. 41*, 124-129.

89. Chaires, J. B., Ren, J., Hamelberg, D., Kumar, A., Pandya, V., Boykin, D. W., and Wilson, W. D. (2004) Structural selectivity of aromatic diamidines, *J Med Chem 47*, 5729-5742.

90. Nguyen, B., Lee, M. P. H., Hamelberg, D., Joubert, A., Bailly, C., Brun, R., Neidle, S., and Wilson, D. W. (2002) Strong binding in the DNA minor groove by an aromatic diamidine with a shape that does not match the curvature of the groove, *J. Am. Chem. Soc. 124*, 13680-13681.

91. Watkins, R. E., Maglich, J. M., Moore, L. B., Wisely, G. B., Noble, S. M., Davis-Searles, P. R., Lambert, M. H., Kliewer, S. A., and Redinbo, M. R. (2003) 2.1 A crystal structure of human PXR in complex with the St. John's wort compound hyperforin, *Biochemistry 42*, 1430-1438.