STRUCTURAL STUDIES ON THE MARR FAMILY, CRYSTAL STRCUTRUES OF OHRR-DNA AND SLYA-INDUCER COMPLEXES

By

Minsun Hong

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 of the degree of
Doctor of Philosophy

September 2005

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

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

Minsun Hong

has been approved.

[redacted]

Mentor

[redacted]

Member

[redacted]

Member

[redacted]

Member

[redacted]

Member

[redacted]

Member

TABLE OF CONTENS

ACKNOWLEDGEMENTS

First of all, I would like to say special thanks to Dr. Dick Brennan for many things including letting me in his lab with a happy working environment, allowing me work on interesting projects and introducing me an exciting world of X-ray crystallography. I always appreciated that Dr. Brennan has been a great mentor (my scientific Dad) and encouraged my thoughts and experiments. I would also like to give very special thanks to Dr. Maria Schumacher for being both a great scientific Mom and friend with her sweet heart. Also I would like to thanks to all of former and current Brennan laboratory peoples including Arthur, Ben, Caleb, Dave, Jason, Joy, Dr. Kate, Kate H, Kevin, Kirsten, Peter, Ryan and Todd for providing pleasant working environments which made me always happy in the lab and being both great co-workers and friends. Specifically, I would love to say thanks to Dr. Kate Newberry for being my crystallography tutor and answering my stupid and bothering questions always with her sweet smiles.

I would like to thank collaborators who allowed me to work with such interesting projects and their helpful discussions; Dr. Helmann and his former graduate student, Mayuree Faungthong who provided the plasmid and cells for OhrR proteins and Dr. Libby for SlyA.

Also I would like to thank my biochemistry classmates, Dave, Margie, Martha S, Martha J, Mellisa, Mee and Lara for being my great friends who had spent my invaluable time together in grad-school.

Finally I am saying "thank you" to my families, my Dad and my Mom, my husband Youngsup, my daughter Kaeun, my parents-in-law, and my brother, who provided unconditional loves and supports.

ABSTRACT

In bacteria, a transcriptional control (from DNA to RNA) is the main target for the regulation of gene expression, which is controlled by the products of regulatory genes, which are often called repressors or activators. These regulatory proteins often bind DNA often through helix-turn-helix motif shared by many DNA binding proteins.

Newly described multiple antibiotic resistance regulatory (MarR) family contains winged helix-turn-helix DNA binding motifs. Members of MarR family regulate the expression of resistance to multiple antibiotics, organic solvents, detergents and oxidative stress agents and to pathogenic factors. At present, biochemical and structural details on MarR family proteins are very limited. No structures of any MarR family member bound to cognate regulatory effectors, DNA, their ligands (or drugs), or chemically modified by its intracellular signal are available.

Structural studies from X-ray crystallography and solution NMR spectroscopy of a variety of gene regulatory proteins, themselves and complexes with cognate DNA or effector molecules have provided their regulatory mechanism at the atomic level. Here X-ray crystallographic studies on OhrR and SlyA were undertaken, in part to address the questions concerning the adaptive responses to reactive oxygen species and transcription regulation of other MarR family proteins. In addition, X-ray structures of OhrR in reduced and bound to *ohrA* promoter DNA shall provide an atomic description of one of the their regulatory mechanisms of Gram-positive and Gram-negative bacteria in response to oxidative stress regulation resulting from alkyl hydroperoxides. X-ray

crystal structures of OhrR in reduced form and DNA bound form and SlyA in ligand-induced form have been determined to aid understanding of regulatory mechanism of OhrR and SlyA as well as other MarR family members.

Chapter 1. Introduction

1. 1 Summary

Unlike animals, microorganisms (and plants) lack the ability to defend themselves from unfavorable conditions of the environment by physical action. Nevertheless, bacteria are endowed with specific means of transmitting alarm (and or emergency) signals. Stress responses are particularly important to microorganisms, because their habitats are subject to continual changes including temperature, osmotic pressure and substrate availability. Bacterial cells respond to alteration in their environment by activating small or large groups of genes under control of a common regulatory protein, a process called global gene regulation that allows cells to cope with specific environmental stresses. Over the years, it has become clear that no one bacterial species is a perfect representative of all others and at a minimum, *Bacillus subtilis* came to be viewed as the paradigm for gram-positive and *Escherichia coli* for any gram-negative bacteria of the microbial world.

In bacteria, transcriptional control (from DNA to RNA) is the main target for the regulation of gene expression harnessed by DNA-dependent RNA polymerase (RNAP). Four subunits of bacterial RNAP ($2\alpha\beta\beta'$) comprise the core enzyme that is capable of basic polymerization activity *in vitro*, but the formation of holoenzyme with sigma factor (σ) is required to initiate efficient transcription from a promoter. There are two major important differences between the mechanisms of transcriptional regulation in bacteria and higher organisms. First is the absence of a nuclear membrane in bacteria, with no barrier between DNA and protein expression. Second is the presence of an operon. A

bacterial operon is a region of DNA that includes genes co-transcribed into the same mRNA plus all of the adjacent *cis*-acting sequences required for transcription of these genes, including the genes' promoter as well as operators and other sequences involved in regulating the transcription of the genes. Because the genes of an operon are all transcribed from the same promoter and use the same regulator sequences, all the genes of an operon can be transcriptionally regulated simultaneously (Snyder and Champness, 1997).

The transcription of a bacterial operon is regulated by the products of regulatory genes, which are often called repressors or activators. These regulatory proteins bind close to or within the operon's promoter and regulate transcription from the promoter. Repressors and activators work in opposite ways. Repressors bind to sites called operators and turn off the promoter, thereby preventing transcription of the genes of the operon. Activators, in contrast, bind to activator sites and turn on the promoter, thereby facilitating transcription of the operon genes (Snyder and Champness, 1997).

These regulatory proteins often bind DNA through a helix-turn-helix (HTH) motif. A region of approximately 7 to 9 amino acids forms an α-helical structure called helix 1, which is separated by about 4 amino acids from another α-helical region of 7 to 9 amino acids called helix 2. The two helices are at approximately right angles to each other, hence the name helix-turn-helix (Steitz et al., 1982). This HTH motif (helix1-turn-helix2) was first identified in the prokaryotic transcription regulatory proteins, cro and the amino-terminal fragment of cI (also called as λ repressor) proteins from

bacteriophage λ and catabolite gene activator (CAP or cAMP (adenosine 3′, 5′-monophosphate) receptor protein (CRP)) from *Escherichia coli* (Anderson et al., 1981; McKay and Steitz, 1981; Pabo and Lewis, 1982; Schultz et al., 1991) (Figure 1. 1A). When these proteins bind to DNA, helix 2 lies in the major groove of the DNA double helix and makes hydrogen bonds with specific bases in the DNA while helix 1 lies crosswise to the DNA and mostly functions to assist in the correct configuration of helix 2. Thus, a DNA binding protein containing a helix-turn-helix motif recognizes and binds to specific regions on the DNA. Moreover, many DNA-binding proteins exist as dimers and bind to invert repeated DNA sequences. In such cases two polypeptides in the dimer are arranged head to tail so that the amino acids in helix 2 of each polypeptide can make the same or similar contacts with the bases in the inverted repeats.

In the absence of structural information, the existence of a helix-turn-helix motif in a protein can often be predicted from the amino acid sequence, since some sequences of amino acids cause the polypeptide to assume an α-helical form and the bent region between the two helices usually contains a glycine. The presence of a helix-turn-helix motif in a protein helps identify it as a DNA-binding protein. However, recent structural work has shown a number of variations on the classical HTH DNA-binding domains in their structures and DNA binding mechanisms. With structural prediction, genetic and mutational studies can give limited knowledge about protein-DNA interactions. But x-ray crystallographic studies are always needed to provide complete understanding of protein-DNA interactions at the atomic level.

In this thesis, x-ray crystallography studies on OhrR and SlyA will be addressed to understand their function and regulatory mechanism in atomic detail. Both proteins belong to a MarR family which contains a newly described winged helix-turn-helix motif. In previous studies, the reactive site on the OhrR protein had been identified and the structural studies described here mapped and allowed a possible induction mechanism to be proposed. At the same time, the structure of OhrR-DNA complex shall address an operator recognition mechanism by OhrR which can be generalized to other MarR family members. The SlyA-citrate complex structure provides the induction mechanism as well as the ligand or inducer binding site of this MarR family member. Such observations shall be compared and provide a general view of the induction mechanism in MarR family members.

1. 2 General stress responses in bacteria

Exposure of bacteria to diverse growth-limiting stresses induces the synthesis of a common set of proteins that protect the cell against future stresses. This general stress response brings about a special physiological state which significantly enhances bacterial survival in the natural environments as well as some pathogenic interactions (Hengge-aronis, 2000).

The general stress response of *E. coli* is characterized by numerous alterations in cellular physiology and morphology that enhance survival by increasing cellular stress resistance, which prevents cellular damage rather than repairs it. This response is triggered by many different stresses including starvation (which results in stationary phase), high osmolarity,

high or low temperature, and acidic pH. These conditions result in the accumulation of σ^s (RpoS), a sigma subunit of RNA polymerase, which acts as the master regulator of this response. Currently more than 50 σ^s-controlled genes that confer stress response have been identified and this number has steadily increased (Hengge-aronis, 2000). σ^s is closely related to the vegetative ("house keeping") sigma factor σ^{70} and they both recognize similar promoter sequences. Thus σ^s is considered a second primary sigma factor rather than an alternative sigma factor (Lonetto et al., 1992). Whereas σ^s is dispensable during exponential growth and, in fact, is hardly present in growing wild-type cells, it is essential in stationary phase (starving) or when cells are stressed (Lange and Hengge-Aronis, 1991). σ^s-controlled promoter regions often feature multiple binding sites for additional regulators such as cAMP-CRP or the histone like proteins H-NS, leucine-responsive regulatory protein (Lrp), integration host factor (IHF) and Fis and these factors play crucial roles in determining sigma factor specificity. Expression of σ^s is mainly controlled posttranscriptionally by the rearrangement of the *rpoS* mRNA secondary structure by *trans*-acting factors such as the RNA binding protein Hfq (also known as host factor I, HF-I), OxyS and DsrA making the RNA-binding site accessible. In addition σ^s levels are regulated by proteolysis, so that in exponentially growing cells, there is measurable synthesis of σ^s but actual σ^s levels are very low. A complex ATP-dependent processive protease (ClpXP) and response regulator (RssB) have been identified as essential components for proteolysis of σ^s (Hengge-aronis, 2000).

In contrast, to *E. coli*, where the general stress response is controlled by σ^s, in *Bacillus subtilis* (another related gram-positive bacteria), this response is controlled by the σ^B, an

alternative σ factor whose activity is regulated posttranslationally by a signal transduction network in which essential protein-protein interactions are governed by serine and threonine phosphorylation (Haldenwang, 1995). It has been estimated that over 200 genes are directly or indirectly under σ^B control which corresponds to at least 5 % of the coding genome (Price, 2000). Loss of σ^B function leads to increased sensitivity to multiple stresses, including acid, heat, osmotic and oxidative stress. The known functions of genes controlled by σ^B are consistent with a role in protecting cellular DNA, protein and lipid against the deleterious effects of these stresses, specifically against the reactive oxygen species generated by unbalanced metabolism to be discussed later. Gene products under σ^B control include the ClpP protease and the ClpC chaperone subunit (heat shock regulon), which are thought to sort, repair or degrade damaged proteins; a Dps homolog, which is thought to bind and protect DNA; and OpuE, which transports proline as a compatible osmoprotectant (Price, 2000).

In addition to the regulon controlled by σ^B, the sporulation process can also be considered a general stress response. However, sporulation is a unidirectional process that consumes substantial time and resources, whereas the general stress response controlled by σ^B is reversible, allowing the cells to readily return to active growth when the stress is removed

1. 3 Oxidative stress responses in bacteria

Unlike eukaryotic cells, bacteria do not contain membrane-bound organelles, so their ability to generate compartments with redox states that are optimized for different cellular processes is limited. All prokaryotes contain in a single compartment, the cytoplasm, so

where the presence of ROS is even more detrimental. Consequently bacteria have evolved sensitive and specific sensors to monitor different redox signals such as the presence or absence of oxygen, cellular redox state or reactive oxygen species (Storz and Imlay, 1999). This adaptive response is coordinated by the action of transcription factors that sense oxidative stress and regulate the expression of appropriate defensive and repair functions (Helmann et al., 2003).

1. 3. 1 Reactive oxygen species (ROS)

Molecular O_2 contains an even number of electrons with two unpaired electrons in two molecular orbitals and is said to be in a triplet ground state. These two electrons have the same spin quantum number, and if O_2 attempts to oxidize anther atom or molecule by accepting a pair of electrons from it, both new electrons must be of parallel spin so as to fit into the vacant spaces of the orbitals. Usually, a pair of electrons in an atomic or molecular orbital would have antiparallel spins. This imposes an important restriction on oxidation by O_2. Molecular O_2 is usually constrained to one electron transfer reaction at a time resulting in formation of superoxide radical ($O_2^{-}\bullet$) and related species, hydrogen peroxide (H_2O_2) and hydroxyl radical (HO•), which contain free radicals (Cadenas, 1989).

A free radical is defined as any molecule or atom which contains one or more unpaired electrons in its outer orbit. The presence of an unpaired electron makes the species highly reactive. These compounds can be formed by the loss of a single electron from a non-radical, by gain of a single electron by a non-radical, or by homolytic fission, in

which a covalent bond is broken and each fragment retains its own electron. A free radical is shown in chemical formulas by a dot, R•. The most noticeable radicals in biological systems are the superoxide anion ($O_2^{\bullet -}$), hydroxyl radical (OH•), nitric oxide (NO•), and the lipid-derived peroxyl radical (ROO•) and alkoxyl radical (RO•). These radicals are termed as reactive oxygen species (ROS) is including other oxygen-related reactive compounds, such as singlet oxygen ($1O^2$), hydrogen peroxide (H_2O_2), and hydroperoxides (ROOH) (Halliwell and Gutteridge, 1989).

All these compounds are detrimental in cells because of their potential to cause free radical reactions (Pryor 1986). When a free radical reacts with a non-radical molecule, the target molecule is converted to a radical, which may further react with another molecule. The primary formation of most of the ROS is the reduction of molecular oxygen with the formation of $O_2^{-\bullet}$. Although the reactivity of $O_2^{-\bullet}$ is relatively low, it can readily initiate free-radical chain reactions. Superoxide undergoes a dismutation to form H_2O_2 spontaneously or enzymatically. In the presence of superoxide dismutases (SOD), the reaction rate is dramatically increased by 10^4 times, indicating its high reactivity.

(I) $O_2^{\bullet -} + O_2^{\bullet -} + 2H^+ \rightarrow H_2O_2 + O_2$: SOD

The product by the dismutation reaction, hydrogen peroxide is more stable than $O_2^{\bullet -}$. However, its permeability through the plasma membrane can cause further radical reaction inside cells if not scavenged by catalase or glutathione peroxidase. The most reactive and potentially harmful radical has been considered to be OH•. Because the lifetime of OH•, it can thus be expected to react at or close to its site of formation

(Gutteridge and Halliwell, 1989). The hydroxyl radical is generated from H_2O_2 through the Fenton reaction catalyzed by the transition metals iron or copper (II):

(II) $Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH\bullet + OH\bullet$

or from $O_2\bullet^-$ and H_2O_2 through the Haber-Weiss reaction catalyzed by iron or copper metal (III)

(III) $O_2\bullet^- + H_2O_2 \rightarrow O_2 + OH\bullet + OH\bullet$: catalyst

Reactive oxygen species are potentially harmful, because they can damage DNA, proteins, lipids and carbohydrates as well as interact with each other and modify cellular biomolecules. Radicals induce DNA strand breaks as well as oxidation of purine and pyrimidine bases, and increase the occurrence of mutations (Cadenas, 1989). In addition, ROS can activate or inactivate proteins by oxidizing sulfhydryl groups and modifying amino acids (Davies, 1987; Davies et al., 1987). In addition, lipid peroxidation is a radical-mediated chain reaction initiated by abstraction of a hydrogen atom from a polyunsaturated lipid and terminated by chain-breaking antioxidants. Excessive peroxidation of membrane lipids disrupts the bilayer arrangement, decreases membrane fluidity, increases membrane permeability, and modifies membrane-bound proteins. However, this damaging effects of ROS can be beneficial to host cells. During phagocytosis, activated inflammatory cells generate ROS through membrane-bound NADPH oxidase - a phenomenon called respiratory burst and a host defense against microbicidal agents (Babior, 1984).

Because ROS [including hydrogen peroxide (H_2O_2), superoxide anion, hydroxyl radical, and organic hydroperoxides (OHP)] can damage proteins, DNA, lipids, and membranes and potentially lead to mutation and cell death, they must be removed rapidly (Gutteridge and Halliwell, 1989; Storz and Imlay, 1999). ROS are generated as byproducts of normal aerobic metabolism and defense responses of both animals and plants (Babior, 1984). To prevent oxidant damage, cells have evolved several antioxidant mechanisms which comprise (Gutteridge and Halliwell, 1989):

(a) agents that catalytically remove free radicals and other 'reactive species'. Examples are the enzymes superoxide dismutase (SOD, $2\ O_2^{\bullet -} + 2H^+ \rightarrow H_2O_2 + O_2$), catalase ($2H_2O_2 \rightarrow 2H_2O + O_2$), peroxidase ($SH_2 + H_2O_2 \rightarrow S^- + 2H_2O$) and "thiol-specific antioxidants'.

(b) Proteins that protect biomolecules against damage including oxidative damage by other mechanisms, e. g. heat shock proteins.

(c) Low-molecular weight agents that scavenge ROS, such as glutathione (GSH, $H_2O_2 + 2GSH \rightarrow GSSG + 2H_2O$).

Some mobile organisms avoid O_2 toxicity by swimming away from regions of high O_2 tension. In several bacteria, including *Samonella typhimurium* and *Escherichia coli*, there is an intracellular redox sensor which measures the redox state of the constituents of the respiratory chain and transmits a signal to the flagellae involved in swimming (Gutteridge and Halliwell, 1989).

1. 3. 2 Oxidative stress response in gram-positive and gram-negative bacteria

The term oxidative stress refers to the situation of a serious imbalance between production of ROS and antioxidant defense. Sie, who introduced the term of Oxidative stress as 'a disturbance in the prooxidant-antioxidant balance in favor of the former, leading to potential damage' (Halliwell and Gutteridge, 1989). In principle, oxidative stress can result from either depletion of antioxidant molecules or increased levels of ROS or both. (Storz and Hengge-Aronis, 2000). To prevent oxidant damage, cells have evolved antioxidant enzymes (AOEs), such as superoxide dismutase, catalase, and glutathione peroxidase, as well as nonenzymatic scavengers. Although damaging, the toxicity of ROS is harnessed for the host defense against micro-organisms and cellular signaling agents (Storz, 2000).

Treatment of bacteria with low levels of oxidants typically results in a classic adaptive response; the treated cells have a greatly enhanced ability to survive subsequent challenge with an otherwise lethal dose of the same oxidant or a related oxidant. This adaptive response is coordinated by the action of transcription factors that sense oxidative stress and regulate the expression of appropriate defensive and repair functions (Storz and Imlay, 1999). Therefore, many bacteria respond to mild oxidative stress by becoming resistant to more severe oxidative stress.

For example, when *E. coli* or *S. typhimurium* are exposed to moderate levels of H_2O_2, the synthesis of about 30 proteins increases and the cells then become resistant to damage by higher levels of H_2O_2. Some of these proteins overlap with those induced in response to heat shock (Storz and Imlay, 1999).

Nine of the proteins induced by H_2O_2 are controlled by one particular gene, called *oxyR*. OxyR is a 34-kDa transcription activator of a member of the LysR family (Christman et al., 1989). Genetic and biochemical studies revealed that activation of OxyR by H_2O_2 is achieved by the formation of an intramolecular disulfide bond between C199 and C208 thiols to induce the *oxyR* regulon (Aslund et al., 1999; Kullik et al., 1995; Zheng et al., 1998). OxyR is deactivated upon reduction by glutaredoxin 1 with the consumption of glutathione (Aslund et al., 1999; Zheng et al., 1998). Both oxidized and reduced OxyR can bind to several gene promoters on DNA, but only the oxidized form can activate transcription. Products of activated genes include hydroperoxidase I (product of the *katG* gene) and alkyl hydroperoxide reductase (*ahpCF* gene). Also induced by OxyR is Dps, a DNA binding protein that has been proposed to protect against oxidative damage, since deletion of the *dps* gene causes *E. coli* to become hypersensitive to H_2O_2.

The OxyR system functions in *E.coli* to maintain a steady-state intracellular level of H_2O_2 of about 0.2 μM over a wide range of growth conditions. Strains lacking *oxyR* showed increased spontaneous mutation rates. OxyR activation is reversed by enzymatic reduction with glutaredoxin 1 (Grx1). The gene encoding Grx1 is regulated by OxyR, thus providing a mechanism for autoregulation (Zheng et al., 1998).

In addition, excess $O_2^{\bullet -}$ generation within *E. coli* leads to increased formation of disulfide bond and will activate OxyR (Demple, 1998). It also activates about ten additional genes, including those encoding MnSOD (*sodA* gene), a DNA-repair enzyme

(endonuclease IV), fumarase C, aconitase A and glucose-6-phosphate dehydrogenase. Fumarase C and aconitase A may be 'back-up' enzymes that are not sensitive to inactivation by $O_2^{\bullet -}$.

In *E. coli*, the SoxRS system employs another major regulon response to oxidative stress. In this system, the oxidative signal to $O_2^{\bullet -}$ is sensed by the protein encoded by the *soxR* gene (Hidalgo et al., 1997). Oxidized SoxR protein activates transcription of the *soxS* gene, whose product binds to the promoters and activates transcription of the genes encoding the proteins. The SoxR protein contains iron-sulfur [2Fe-2S] clusters active in promoting *soxS* transcription only in their oxidized $[2Fe\text{-}2S]^{2+}$ state. In addition, activation of the SoxRS system also increases resistance to several antibiotics, due to decreased synthesis of a membrane protein, OmpF.

Moreover, a detrimental characteristic of some metal ions, particularly Fe (II), is ability to react with hydrogen peroxide (H_2O_2) to produce the damaging hydroxyl radical (·OH). It is therefore necessary that intracellular levels of both reactive oxygen species (ROS) and Fe (II) be tightly regulated. *E. coli* has a high affinity iron-uptake system, comprising proteins encoded by about 30 genes. Their transcription is increased when the cell iron content is low. This system is regulated by the *fur* (ferric uptake regulator) gene. If the cells have enough iron, the product of this gene binds Fe^{2+} and the complex binds to a DNA base sequence (the iron box) found in the promoter regions of many genes whose expression is regulated by iron. As a result transcription is blocked. Genes

with iron boxes include not only the iron-uptake systems but also those encoding the enzymes hydroperoxidase I and II and MnSOD.

Bacillus subtilis displays a complex adaptive response to low levels of various peroxides, controlled by σ^B (Helmann et al., 2003), PerR (Herbig and Helmann, 2001) and OhrR (Fuangthong et al., 2001).

Protection of starved cells against oxidative stress was the first role attributed to the σ^B regulon, and sensitivity to this stress remains one of the strongest phenotype of a *sigB* null mutant. Loss of σ^B function leads to decreased resistance to organic hydroperoxide because products of σ^B regulon specifically prevent or repair the damage caused by oxidative stresses. Theses include the KatB (Engelmann et al., 1995) and KatX (Petersohn et al., 1999) catalses and the putative YdbD manganese-containing catalase (Price et al., 2001), which destroys peroxides and related damaging compounds, and the TrxA thioredoxin which helps maintain thiol-disulfide balance (Scharf et al., 1998). Due to the absence of a glutathione system, TrxA function is essential in *B. subtilis* (Scharf et al., 1998).

A second protein in the *B. subtilis* oxidative stress response, PerR (MW ~21 kDa), is a member of the ferric uptake repressor (Fur) family of metal-dependent DNA-binding proteins (Bsat et al., 1998). PerR has a helix-turn-helix DNA binding motif in its N-terminous and contains two CXXC metal binding motifs in the C-terminal putative metal-binding domain. (Bsat et al., 1998; Fuangthong et al., 2002). The PerR regulon includes

the DNA-binding protein (mrgA), a major vegetative catalase (*katA*), alkyl hydroperoxide reductase (*ahpCF*), a heme biosynthesis operon (*hemAXCDBL*), a zinc uptake system (*zosA*), *fur* and *perR* itself (Herbig and Helmann, 2002). However not all components of the PerR regulon are inducible by peroxide such as *fur* and *perR*. The metal selectivity of PerR varies. The repression of both *mrgA* and *katA* by PerR can be elicited by iron, and both genes are strongly induced by H_2O_2. In contrast, repression of both *fur* and *perR* appears to be selective to Mn (II) and neither gene can be strongly induced by H_2O_2. However, the *zosA* gene is also selectively repressed by Mn (II), but this gene is strongly induced by H_2O_2 (Fuangthong et al., 2002). Therefore, the identity of the regulatory metal ion cofactor influences the sensitivity of PerR to oxidants: the iron-containing form is quite sensitive to H_2O_2, while addition of Mn (II) reduces this sensitivity (Herbig and Helmann, 2001). However, the atomic details of oxidation by peroxide and reactivation of oxidized PerR are unclear.

The PerR regulon of *B. subtilis* resembles the enteric peroxide stress regulon controlled by OxyR (Storz, 2000). Both systems consist of a group of genes induced by peroxide stress. However, the *B. subtilis* system is under the negative control of PerR and this regulation has an intimate relationship with the level of intracellular metal ions. In contrast, OxyR is an activator of peroxide stress genes and is not a metalloprotein. Their activation mechanisms are indeed quite different. In *E. coli*, the H_2O_2-catalyzed disulfide bond formation activates OxyR (Zheng et al., 1998).

As described in this chapter, the detoxification of H_2O_2 and superoxide radical is well understood and their regulatory mechanisms are reasonably well defined (reviewed in (Paget and Buttner, 2003; Storz and Imlay, 1999). However, those of organic hydroperoxides are less well understood. These peroxides threaten cell integrity as they can attack long chain fatty acids and other nonpolar components of the cell membrane.

The organic hydroperoxide resistance protein regulator (OhrR) (reviewed in (Paget and Buttner, 2003), a member of the MarR family, is a novel organic peroxide-sensing transcription factor that controls expression of organic hydroperoxide resistance (ohr) protein, by binding to *ohrA* promoter. Ohr, which is present in both Gram-positive and Gram-negative bacteria, is essential for the oxidative stress response against reactive oxygen species (ROS) (Atichartpongkul et al., 2001; Mongkolsuk et al., 1998) by sensing the presence of organic peroxide (Atichartpongkul et al., 2001; Fuangthong et al., 2001; Ochsner et al., 2001; Shea and Mulks, 2002). Because ROS, including hydrogen peroxide, superoxide anion, hydroxyl radical, and organic hydroperoxides, can damage proteins, DNA, lipids, and membranes and eventually lead to cell death, they must be removed rapidly (Storz and Imlay, 1999). ROS are generated as byproducts of normal aerobic respiration and as reaction intermediates during host defense responses following the release of the lysosomal contents within inflammatory cells and the neutrophil oxidative burst (Atichartpongkul et al., 2001; Shea and Mulks, 2002). Recent structural and biochemical studies of Ohr have suggested that the Ohr protein contains alkyl hydroperoxide reductase activity for organic hydroperoxide (OHP) detoxification (Cussiol et al., 2003; Lesniak et al., 2002; Mongkolsuk et al., 1998).

B. subtilis OhrR is a homodimer of 147 amino acid residues per subunit (mol. Wt 17 kDa). A unique and conserved cysteine residue, Cys15, is oxidized to Cys-sulfenic acid (CysSOH) by organic hydroperoxides, such as *tert*-butyl hydroperoxide and cumen hydroperoxide, which result in the derepression of the ohr organic peroxide resistance gene, *ohrA* (Fuangthong and Helmann, 2002). Because of its high in vivo reactivity, the CysSOH acid may take higher oxidation states by Cys-sulinic acid ($CysSO_2H$). In OxyR, described as the global regulator of peroxide stress in *E.coli*, the sulfenic acid modification of residue Cys199 has been proposed to be an intermediate in the formation of a disulfide bond between residues Cys199 and Cys208 (Choi et al., 2001; Zheng et al., 1998). However, the formation of an intradimeric disulfide bond has not been seen in OhrR (Fuangthong and Helmann, 2002).

Thus, there are at least three regulatory mechanisms for sensing peroxide stress observed in Bacteria: regulation by thiol-disulfide exchange by OxyR through a formation of disulfide bond, metal-mediated peroxide-sensing by PerR and a cysteine oxidation by OhrR through the formation sulfenic acid of which further oxidation state is currently unknown.

1. 4 MarR family proteins

Structural studies from X-ray crystallography and solution NMR spectroscopy of a variety of gene regulatory proteins and their complexes with cognate DNA or effector molecules have provided information on their regulatory mechanism at the atomic level.

Current structural analysis has revealed that most prokaryotic transcription factors are homodimers that bind to palindromic or pseudopalindromic operator DNA in a sequence-specific manner. Most of these transcription factors employ one of three major DNA-binding motifs to recognize their own operator DNA sites: the helix-turn-helix (HTH), the winged helix-turn-helix (wHTH) and the β ribbon (Brennan, 1993; Wintjens and Rooman, 1996). One of the best described, the HTH motif, is composed of a two to three turn helix, turn and a four turn "recognition" helix (Brennan, 1993). This HTH motif was first identified in prokaryotic transcription regulatory proteins, cro and the amino-terminal fragment of the cI (also called as λ repressor) proteins from bacteriophage λ and catabolite gene activator (CAP or cAMP (adenosine 3′, 5′-monophosphate) receptor protein (CRP)) from *Escherichia coli* (Figure 1. 1A) (Anderson et al., 1981; McKay and Steitz, 1981; Pabo and Lewis, 1982; Schultz et al., 1991), DNA binding motifs of which were completely α-helical. Since then, subsequent structural works have shown a number of variations of the classical HTH DNA-binding domains. Most notable are the differences in topology (Wintjens and Rooman, 1996) observed in several α+β domains where β strands interrupt, precede or follow the canonical HTH motifs helices involved in DNA-binding. An example of this α+β domains is the winged helix-turn-helix proteins.

This thesis describes crystallographic studies on OhrR and SlyA that were undertaken, in part to address questions concerning the MarR member structures. Members of the multiple antibiotic resistance regulatory (MarR) family regulate the expression of

resistance to multiple antibiotics, organic solvents, detergents, oxidative stress agents and pathogenic factors. At present, biochemical and structural details on MarR family proteins are very limited. No structures of any MarR family member bound to cognate regulatory effectors, DNA, their ligands (or drugs), or chemically modified by its intracellular signal are available. The crystal structure determination of MarR is the first step in understanding the structural basis for the function of the MarR family of proteins. The structure also advances efforts aimed at identifying small molecules that could modulate the function of these proteins and, thus serve as novel therapeutics for a number of medically important pathogens.

1. 4. 1 Definition and members of MarR family

MarR was originally identified as a component of the *Escherichia coli mar*RAB locus that negatively regulates expression of this operon (Alekshun and Levy, 1997). The product of *marA* is a transcription factor that autoactivates expression of the *marRAB* operon and regulates the expression of a global network of >60 chromosomal genes (Barbosa and Levy, 2000; Martin et al., 1996). Later the definition of MarR family protein has been extended to define proteins of a family that controls an assortment of biological functions, including resistance to multiple antibiotics, organic solvents, detergents and oxidative stress agents, thereby collectively termed the multiple antibiotic resistance (Mar) phenotype (Alekshun and Levy, 1999). These proteins also regulate the synthesis of pathogenic factors in microbes that infect humans and plants (Miller and Sulavik, 1996). As members of the MarR family, MarR, MexR, SlyA and OhrR have been characterized and will be discussed.

MexR from *Pseudomonas aeruginosa* is a negative regulator of the *mexRAB-oprM* multidrug efflux operon (Poole et al., 1996; Srikumar et al., 2000). *P. aeruginosa* is an opportunistic human pathogen due to its multidrug resistance (Hancock, 1998). Mutations is *mexR* lead to overexpression of the *mexAB-oprM* operon, resulting in increased resistance to multiple antimicrobals, including fluoroquindlones, B-lactams, tetracycline, macrolides, chloramphenicol, novobiocin, trimethoprime, and sulfonamides (Nikaido, 1998; Srikumar et al., 2000). The atomic coordinates and structure factors for MexR have been deposited under the code 1LNW.

SlyA is a transcription regulator present in *Escherichia coli*, *Salmonella enterica*, and other bacteria belonging to the *Enterobacteriaceae*. *S. typhimurium* SlyA is a 144-amino acid protein (MW ~16.5 kDa) and required for both the virulence in mice and the survival of salmonellae in murine macrophages (Libby et al., 1994). In addition, SlyA has been shown to play a role in oxidative stress response in the respiratory burst during infection of the host (Buchmeier et al., 1997b). Recent two-dimensional polyacrylamide gel eletrophoresis (2D-PAGE) studies, aided by mass spectrometry in *E. coli* and *Salmonella* serovar Typhimurium, have shown that SlyA is involved in the transcriptional regulation of other stress response proteins. Such findings suggest a crucial role of SlyA for intracellular survival and replication of those bacteria in phagocytic host cells (Spory et al., 2002).

B. subtilis OhrR is a homodimer of 147 amino acid residues (mol. Wt 17 kDa). A unique and conserved cysteine residue, Cys15, is oxidized to Cys-sulfenic acid (CysSOH) by organic hydroperoxides, such as *tert*-butyl hydroperoxide and cumen hydroperoxide, which result in the derepression of the ohr organic peroxide resistance gene, *ohrA* (Fuangthong and Helmann, 2002). Because of its high reactivity, the formation of higher oxidation states of CysSOH acid has been proposed. However, the formation of an intradimeric disulfide bond has not been observed in OhrR (Fuangthong and Helmann, 2002).

1. 4. 2 Structures of MarR family proteins

MarR family proteins belong to winged helix turn helix (wHTH, also called winged helix proteins) DNA-binding superfamily which constitute a subfamily within the large ensemble of helix-turn-helix proteins. Topologically, the winged helix motif is a compact α/β structure consisting of wing, three α helixes (H1, H2, H3) and three β strands (S1, S2 and S3), arranged in order H1-S1-H2-H3-S2-W1-S3. Recently this MarR family has been sub-divided into true MarR and MarR-like families, because the latter contain the wHTH DNA binding motif of MarR but different features in other respects.

Despite low primary sequence similarity amongst MarR family proteins, less than 20%, the members of this family show significant structural homology. X-ray crystal structures of MarR family proteins, including MarR, MexR, SarR and SlyA-like proteins, have revealed a common triangular shape with two winged helix-turn-helix (HTH) DNA-binding motifs (Figure 1. 2A). These proteins have been shown to consist of two

domains: and NH2-terminal dimerization domain and the COOH-terminal DNA binding domain.

However, superposition of several MarR family proteins indicates a conformational difference in the dimerization helices (α1, α 5 and α 6, Figures 1. 1B and 1C). In the MexR crystal structure, four MexR dimer were present in the crystallographic asymmetric unit and significantly different conformations were observed in helices α1, α5 and α6, loops between α1 and α2, α5 and α6 and β2 and β3 in the various subunit. This suggests that natural structural flexibility can readily cope with the different specificities of their cognate effectors and operator DNA adaptation and their modes of regulation.

One more thing to note about the X-ray crystal structures determination of MarR family proteins, all MarR family protein structures, except SarR, have determined by *de novo* methods. SarR was solved in a maltose binding protein (MBP) conjugated form by molecular replacement where the structure of MBP was used as a template. Considering their relatively small sizes, the observed conformational flexibilities in their tertiary folds are significant enough to make molecular replacement unsuccessful.

1. 4. 3 Operator DNA binding

The manner in which MarR family proteins binds to DNA remains to be determined. Without any available MarR family protein-operator DNA structure, mutational and biochemical studies on MarR and MexR have given some clues regarding operator DNA

binding by MarR family proteins. Footprinting experiments suggested that MarR binds as a dimer at two separate but very similar sites in *marO*, protecting ~21 base pairs of DNA on both strands at a single site and not bending its target (Martin et al., 1996; Martin and Rosner, 1995). For MarR to bind as a dimer, with each winged HTH DNA binding domain must contact one half-site on B-DNA. Geometric constraints suggests only a few possible modes of binding. First is the binding of a single dimer to one operator DNA which would require reorientation of the DNA binding lobes so that each could reach one half-site. This would be analogous to the binding of an E2F-DP heterodimer (a eukaryotic transcription factor in which each subunit also has a winged-helix DNA binding domain) to its cognate binding site (Figure 1. 1B) (Zheng et al., 1999). A second possible way would involve the binding of two dimers, on opposite faces of the double helix, to a single MarR binding site, analogous to the binding of DtxR (a bacterial protein with a winged-helix DNA binding domain) to its target (Figure 1. 1C). However, the DtxR half-sites are on the same face of the DNA helix (White et al., 1998).

With respect to DNA recognition mechanisms, two possible modes previously observed in other winged-helix family of DNA-binding proteins could be applied in MarR-DNA recognition. First, members such as hepatocyte nuclear factor-3γ (HNF-3γ) use the recognition helix (H3) of the motif as the primary determinant for DNA-protein interactions in the major groove and a wing regions (W1) to form minor groove and phosphodiester backbone nucleoprotein contacts (Clark et al., 1993). Second, other proteins such as a human regulatory factor X1 (hRFX1) use W1 to interact with the major

groove, and the H3 helix makes only a single minor groove contact (Figure 1. 1D) (Gajiwala et al., 2000).

In MarR, mutations in both α4 (H3) and W1 affect the DNA binding activity (Alekshun et al., 2000). For example, mutations in α4, including an R73C change, abolish MarR DNA binding activity (Alekshun et al., 2000). Also, R94C and R91C mutations which are located at the tip of W1, reduced DNA-binding affinity in MarR and MexR respectively (Alekshun et al., 2000; Saito et al., 2003). Furthermore, the presence of several positively charged residues shown in electrostatic potential maps indicate good docking sites for the negatively charged DNA backbone. These data suggest that both the recognition helix and the wing participate in DNA recognition.

In the MarR structure, the juxtaposition of the DNA-binding lobes does not allow for modeling of the whole dimer onto a B-DNA representation of the operator. In MexR structure studies, Strynadka's group has proposed a model of the MexR-DNA complex (Lim et al., 2002). As mentioned earlier, there were four MexR in the asymmertic unit. One of the dimers shows the largest spacing between DNA binding domains with a Cα-Cα distance from the recognition helix (α4) of ~29 Å, which is close to the 34 Å spacing between major groove in linear B-DNA. Given the observed conformational flexibility of the dimerization interface in MexR as well as MarR family proteins, the wings were positioned to make minor groove interactions with further increased spacing between recognition helices.

However, in the absence of protein-DNA cocrystal structures, the precise mechanism by which these mutations affect the DNA binding activity of the protein is uncertain.

1. 4. 4 Effector molecule binding

In the MarR structure, two potential drug-binding sites have been proposed because two salicylate molecules per monomer were found (Figure 1. 3A). Sodium salicylate, at millimolar concentrations, is known to inhibit MarR activity (Alekshun and Levy, 1999) and is routinely used as a model inhibitor of MarR to induce MarA expression in *E. coli* and *Salmonella typhimurium* (Cohen et al., 1993; Sulavik et al., 1997), thereby conferring a Mar phenotype. These sites are on the surface of the molecule on either side of the proposed DNA-bindng helix α4 (H3) which could explain their inhibition of DNA binding by MarR. The SAL-A site is located in a crevice between the recognition helix and the β-sheet (wing) while the SAL-B is located on the opposite site of α4 (Figures 1. 3A and 3B). Whether one or both of the salicylate sites are physiologically relevant or are an artifact of the crystallization conditions are unclear because of the very high concentration of sodium salicylate (250 mM) used in MarR structure determination to provide stable crystals (Alekshun et al., 2001). In the absence of a structure of the ligand-free form of MarR, the effect of occupancy of these sites on protein conformation is is also uncertain.

It has been proposed that the effector for MexR might be a peptide signaling molecule. However, again no true ligand or ligand binding sites for MexR as well as SlyA have been identified.

The major merit for structural studies on OhrR is that its reactive site and its effector molecules have been identified and very well characterized biochemically and genetically by Helmann's group (Fuangthong et al., 2001; Fuangthong and Helmann, 2002).

1. 4. 5 Summary

Members of the multiple antibiotic resistance regulatory (MarR) family regulate the expression of resistance to multiple antibiotics, organic solvents, detergents, oxidative stress agents and pathogenic factors. At present, biochemical and structural details on MarR family proteins are very limited. No structures of any MarR family member bound to cognate regulatory effectors, DNA, their ligands (or drugs), or chemically modified by its intracellular signal are available. In this thesis, crystallographic studies on OhrR and SlyA were undertaken, in part to address the questions concerning the adaptive responses and transcription regulation of other MarR family proteins. At the same time, X-ray structures of OhrR in reduced and bound to *ohrA* promoter DNA will provide an atomic description of one of the regulatory mechanisms of Gram-positive and Gram-negative bacteria in response to oxidative stress regulation resulting from alkyl hydroperoxides.

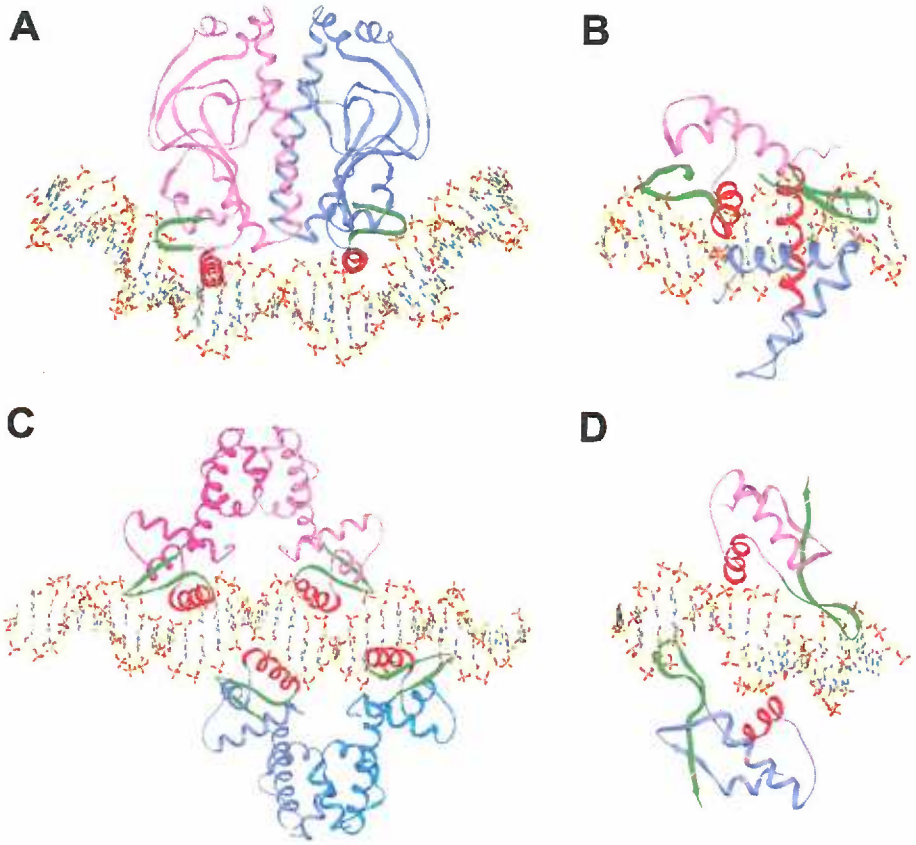


Figure 1. 1 Structures of winged helix-turn-helix motif proteins. Structures of **A**, CAP-DNA complex (PDB accession code, 1CGP) (Schultz et al., 1991), **B,** E2F-DP and DNA complex (1CF7) (Zheng et al., 1999), **C**, DtxR-*tox* operator DNA complex (1DDN) (White et al., 1998) and **D**, human regulatory factor X (1DP7) (Gajiwala et al., 2000). Proteins are shown as ribbon diagrams and recognition helices and wings are colored in red and green respectively. DNA is shown as a stick model with carbons colored yellow; nitrogens, blue; oxygens, red and phosphorus, orange.

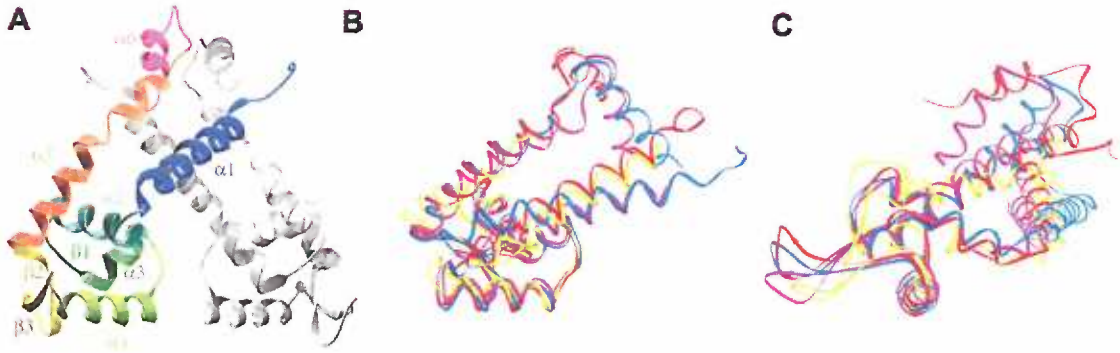


Figure 1. 2 Structures of MarR family proteins. **A**, Overall structure of the MarR dimer, (adapted from Alekshun *et al.*, 2001). **B**, Overlays of the structures of MarR (PDB ID code 1JGS), MexR (1LNW), SlyA-like (1LJ9), and SarR (1HSJ) monomers, colored in red, blue, yellow and purple, respectively. The monomers were superimposed using the main chain atoms of the recognition helix, α4, in A. **C**, The overlays rotated by 90°.

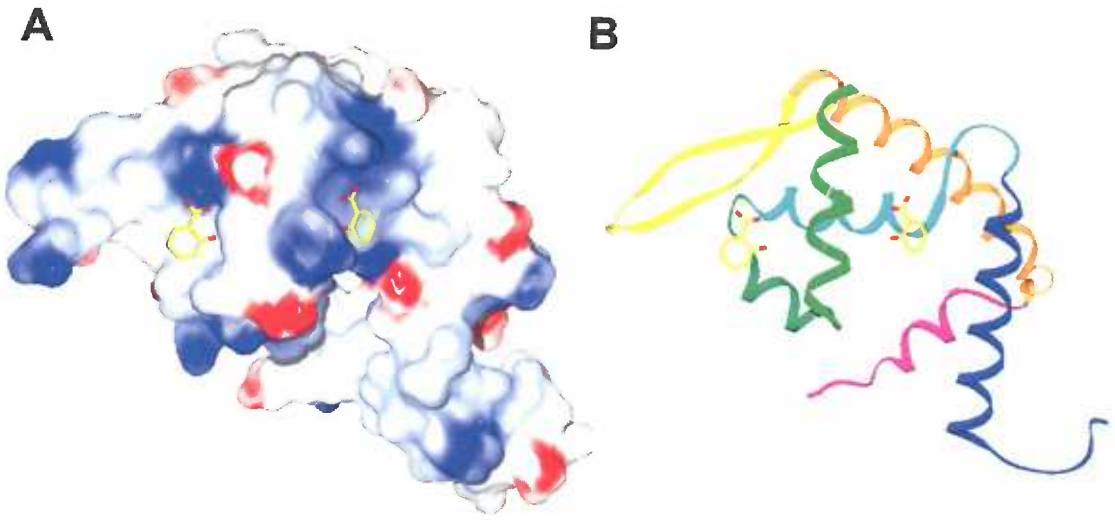


Figure 1. 3 Salicylate binding pockets in MarR structure. **A**, Electrostatic surface potential map of MarR dimer, negatively and positively charged residues colored red and blue, respectively. Salicylate molecules are shown as a stick model. Carbon and oxygen atoms are colored as yellow and red, respectively. **B**, Ribbon diagram of MarR dimer with salicylate molecules.

Chapter 2. Macromolecular structure determination by X-ray crystallography

2. 1 Introduction

Structural studies of macromolecules are performed by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy techniques. Macromolecular X-ray crystallography, which provides the atomic description of macromolecules, will be addressed in this thesis.

Macromolecular X-ray crystallography is a method used to obtain a structural model from X-ray diffraction data from its crystals. Based on the diffraction pattern obtained from X-ray scattering off the periodic assembly of molecules or atoms in the crystal, the electron density can be reconstructed from which a structural model of a macromolecule can be derived. However there is a loss of phase information, which results from measuring X-ray diffraction data from a three-dimensional crystal to a two-dimensional image plate (detector) to present a complete reconstruction of the molecular structure. This is referred to as the " Phase problem" which is the largest stumbling block in crystallography. In this chapter, basic concepts in macromolecular X-ray crystallography and one of the most commonly used methods to solve the phase problem, multiple wavelength anomalous dispersion (MAD), will be described.

2. 2 Basic concepts in X-ray crystallography

X-rays have the proper wavelength (in the Ångström range, 1 Å = 10^{-8} m) for determining macromolecular structures. In order for an object to diffract light, the wavelength must be no larger than the object. The X-rays used for X-ray crystallography usually have wavelengths between 0.5 Å and 1.6 Å. This range gives an appropriate wavelength distance necessary to make an image on the atomic scale. Atoms that are covalently bonded to each other are 1-2 Å apart. Strong polar interactions and hydrogen bonds have atomic distances of 2.5-3.5 Å. In practice, a laboratory (in-house) X-ray generator usually produces X-ray radiation from a copper anode (CuKα, 1.54 Å wavelength). Since the 1960s, synchrotrons have become available which are much more intense X-ray sources. Synchrotron sources produce a continuous X-ray spectrum, from 0.5 Å to 1.6 Å (Blow, 2002b). Additionally the tunable property of synchrotron X-ray radiation greatly improved macromolecular X-ray crystallography by allowing scientists to utilize the anomalous scattering properties of certain atoms at different wavelengths. This is the basis of MAD experiments which will be discussed later in this chapter.

Scattering of X-rays are due to their interaction with the electrons of an atom. "Heavier" are defined as those having more electrons which scatter X-rays more strongly. Hence the result of an X-ray structure experiment represents distributions of electron density. Scattered X-rays have both an amplitude and a phase. The amplitude is magnitude or "intensity" of the wave and the phase is the position of a wave's maximum relative to an origin. Scattering from crystals is often called "**Diffraction**". In order to measure the

scattered X-rays by conventional detectors, a repetitive, ordered array of the scattering objects is necessary to intensify the X-ray intensities. **Crystals** are such an ordered three-dimensional assembly of macromolecules that repeat periodically in all three directions. A regular repeating block, that defines the symmetry of the crystal, is termed the **unit cell** of the crystal, which is the smallest repeating unit. The "**asymmetric unit**" is the unique repeating unit which can be rotated and translated to generate one "unit cell" using the symmetry operators allowed by the crystallographic symmetry. The asymmetric unit can be one molecule or one subunit of oligomeric macromolecules. The relationship between the repeating units in the crystal is described as the "**Symmetry**". The symmetry operators describe how the asymmetric units are arranged with respect to each other in the unit cell by rotation and translation. The shape of any unit cell is described by six parameters: three vectors a, b and c and their interaxial angles α, β and γ. In protein crystallography, the axial lengths are expressed in terms of Å and the interaxial angles are expressed in terms of ° (degrees) (Giacovazzo et al., 1992) (Figure 2.1). These six cell parameters are used to characterize each of the seven basic crystal systems (triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic). The triclinic system has the lowest symmetry in which all six parameters are variable due to its lack of internal symmetry. The other systems, which contain internal symmetry have less independent parameters. In the tetragonal system, which is represented in this thesis by the reduced OhrR crystals, $a = b \neq c$ and $\alpha = \beta = \gamma = 90°$. In the trigonal and monoclinic systems observed in the OhrR-*ohrA* complex crystals, $a = b \neq c$, $\alpha = \beta = 90°$ and $\gamma = 120°$, and $a \neq b \neq c$, $\alpha = \gamma = 90°$ and $\beta > 90°$, respectively. For the monoclinic system, the

b-axis is conventionally taken to be the axis which is perpendicular to both the a- and c-axes.

A three-dimensional unit cell can be also considered as a building block in the crystal lattice. Each of the seven crystal systems addressed above contain lattice points at the corners of the unit cell: that is, there is the equivalent of one lattice point per unit cell. Such lattices are termed primitive and are designated by the letter "P" preceding the symmetry symbols in crystallography. In Bravais' study of lattices, he discovered some lattices that were more complex but still conformed to the symmetry of one of the seven crystal systems. These lattices are termed as nonprimitive lattices. These nonprimitive lattices contain two or more lattice points per unit cell and can be most simply viewed as the combination of a primitive lattice with one or more offset identical copies of itself. A nonprimitive lattive with a pair of lattice points centered on opposite faces of the unit cells is designated by A, B or C depending on whether the bc, ac or ab faces are centered. If there is a lattice point at the body center of a unit cell it is designated by I (inner) and if all faces of the unit cells have lattice points at their centers the designation is F (Stout and Jensen, 1989).

In any array of lattice points, it is always possible to choose a primitive triclinic cell regardless of the symmetry present. But to disregard that symmetry would be to neglect the simplification and advantages provided in crystal symmetry, since higher symmetry will reduce the number of unique parameters to be determined. Therefore, in selecting the unit cell, a cardinal rule is choose in such a way that it conforms to the highest

possible symmetry present in the crystal due to symmetry elements (*e. g.*, rotary axes, mirrors, centers of symmetry and rotary inversion axes).

Crystals can be further devided in terms of the group of symmetry operators relating their faces. Known as point groups, each of them represents one of the possible unique combinations of crystallographic symmetry elements. There are 32 point groups with the 14 Bravis lattices and combining these gives 230 unique arrangements of points in space, called a "space group". The 230 space groups describe the ways in which identical objects can be arranged in an infinite lattice. However, due to the presence of asymmetry in the biological amino acids and that fact that only L-isomers occur, we can exclude space groups with mirror planes and inversion centers, thus reducing the number of possible space groups for protein crystallography to 65 (Drenth, 1999; Stout and Jensen, 1989).

X-ray diffraction occurs where X-rays are scattered from an ordered array of molecules. Each point in the lattice causes scattering but a diffracted beam only arises if the scattered X-rays from each unit are all in phase. Otherwise the scattering from one unit is cancelled out by another. This phenomenon is described by **Bragg's law**, in a crystal the two planes will scatter in phase if this path difference, $2d \sin \theta$, is a whole number of wavelengths, λ, of the X-rays (n):

$$n\lambda = 2d \sin \theta \qquad (2.1)$$

where d is the spacing between the planes and θ is the glancing angle (Figure 2.2) (Drenth, 1999). Bragg's law assumes that phase differences ($n\lambda$) between scattered rays

at different points depend only on path length differences ($2d \sin \theta$). There is no intrinsic phase change between the incident and scattered rays in normal X-ray scattering (Powell, 2005). However, this is not always the case, especially for anomalous scattering which will be discussed later in this chapter.

In addition, Bragg's idea regards each diffracted X-ray by a crystal as a **reflection** in a set of lattice planes. This set of planes is defined by three indices h, k and l, called Miller indices. The planes cut the x-axis in a/h, y in b/k and z in c/l pieces as shown in (Figure 2.1). In crystallography the real lattice is defined in x, y and z coordinates, and the reciprocal lattice, in h, k and l coordinates, are used to explain diffraction as a reflection. The reciprocal lattice vectors are perpendicular to the real lattice planes from which they are derived. There is an inverse relationship between the size of the reciprocal lattice planes and those of the real space lattice. Thus large unit cells result in a very closely spaced reciprocal lattice and small unit cells result in a reciprocal lattice with large spacing. Equation (2.1) can be modified to a more standard way of expressing it, in which particular set of planes in any reflecting position d_{hkl} for any value of n,

$$\lambda = 2\, d_{hkl} \sin \theta \qquad (2.2)$$

where d_{hkl} is the minimum interplanar spacing of the real space lattice for the corresponding reciprocal lattice point (reflection) that is being measured. This is called the "resolution limit" or "**Resolution**". It is directly related to the optical definition in which it is the minimum distance that two objects can be apart and still be seen as two separate objects. Resolution is normally quoted in Å in X-ray crystallography (Cambell, 2005).

In an X-ray diffraction experiment the direction of the diffracted beams depends on two factors: the unit cell lengths in the crystal from which the unit cell distances in the reciprocal lattice are derived and the X-ray wavelength. Figure 2. 3 shows that such diffraction conditions are determined not only by the reciprocal lattice, but also by the radius of Ewald's sphere. When the Bragg reflection angle (an angle between incident beam and crystal lattice) is θ, the total angle of diffraction is 2θ in Figure 2. 3. If one places an intersection of the incident ray with a plane at the center of a sphere with its radius of $1/\lambda$:

$$\sin\theta = \frac{1/2\,(OR)}{(1/\lambda)} \qquad \lambda = 2\sin\theta\,/OR \qquad (2.\,3)$$

Bragg's law can only be satisfied (equation 2. 1) when OR = $1/d$ and $n = 1$. This observation was noted by Ewald in 1921 and introduced a new way of looking at X-ray diffraction as a reflection in a reciprocal lattice *hkl*. This establishes that whenever a reciprocal lattice point coincides with a circle constructed with a radius of $1/\lambda$, which is called the Ewald's sphere or the sphere of reflection, reflection occurs (Blow, 2002b).

Because of its wave and periodic natures, X-ray scattering can be described by a "**Fourier series**" and each reflection given as a "**Structure factor** f". f is a complex number (a vector) containing the amplitude and the phase of the reflection. Amplitude is magnitude or "intensity" of waves. Phase is the position of a wave's maximum relative to an origin. Knowledge of the direction of the diffracted X-rays can determine the dimensions of the unit cell, but, the information of an interest for crystallography is the

contents of the unit cell. In a crystal, electrons in every atom contribute to the amplitude and phase of each individual reflection, not to varying extents in a given crystal system and can be determined by its scattering factor (f) and its position in the unit cell ($hx + ky + lz$). This is called the structure factor.

$$F_{hkl} = \Sigma_j f_j \exp \{2\pi i (hx_j + ky_j + lz_j)\} \quad (2.4)$$

where f_j is the scattering factor from the j^{th} atom, i is the imaginary square root of -1 and x_j, y_j and z_j are the fractional coordinates of the j^{th} atom. Since X-ray scattering is due to electrons, the equation (2. 3) can be also written with f expressed as a function of electron density ρ.

$$F_{hkl} = \int_x \int_y \int_z \rho (x, y, z) \exp \{2\pi i (hx_j + ky_j + lz_j)\} \, dxdydz \quad (2.5)$$

$$F_{hkl} = \int_V \rho (x, y, z) \exp \{2\pi i (hx_j + ky_j + lz_j)\} \, dV \quad (2.6)$$

where the equation (2. 6) is written in terms of the volume (V) of unit cell. Thus the Fourier transform represents the mathematical relationship between the electron density and the diffraction by X-rays. By inverse Fourier transform, the electron density can be calculated:

$$\rho (x, y, z) = (1/V) \Sigma_h \Sigma_k \Sigma_l \left| F(hkl) \right| \exp \{-2\pi i (hx_j + ky_j + lz_j) + i \alpha_{hkl}\} \quad (2.7)$$

where the electron density at each point $\rho (x, y, z)$ is made up of the sum of all the reflection amplitudes $\left| F(hkl) \right|$ and phases α_{hkl}. $\left| F(hkl) \right|$ is also called the structure factor amplitude which can be obtained from the measurements of reflection intensities, $I(hkl)$ by:

$$I(hkl) = \alpha \left| F(hkl) \right|^2 \quad (2.8)$$

However, α_{hkl}, the phases, which is necessary to calculate ρ (x, y, z), cannot be determined directly from the reflections. The correct phases are necessary in order to calculate the electron density. It has been noted that the phase information is more important in electron density calculation the amplitude information by Hauptman and Karle in 1962.

The phase problem that results when only amplitudes are measured, can be solved by three common methods in macromolecular X-ray crystallography: molecular replacement (MR), multiple isomorphous replacement (MIR) and multiple anomalous dispersion (MAD).

The principle of MR is to use a homologous proteins as a search model, which is sufficiently close to the unknown structure that initial phases can be derived by superimposing a known homologous structure onto the observed diffraction data for the unknown structure. Theoretically, one X-ray diffraction data set is required to solve a structure by MR while the *de novo* phasing techniques, MIR and MAD, need more than two or more data sets.

MIR is a primary and the traditional method for determining the initial phases from a new structure by taking advantage of the fact that heavier elements (*e. g.* metals) diffract X-rays more strongly than typical atoms in macromolecular crystals (*e. g.* carbon, nitrogen and oxygen). Theses phases are derived from multiple (usually two or more) data sets collected on crystals into which heavy atoms have been soaked or co-crystallized.

Comparing differences between those data sets allows one to determine the phases of the specifically located heavy atoms and therefore to define the phases in the entire unit cell. In practice, differences in unit cell parameters between heavy atom derivatized and native crystals ("nonisomorphism") can generate serious problems in this method of phase determination.

MAD (Hendrickson, 1991) is newer method for deriving initial phases. This is done by measuring diffraction data at several different wavelengths near the absorption edge of a heavy-atom from a single crystal. The anomalous signal that results from this can give very accurate phases since only one crystal is used.

Two methods, MR and MAD have been applied in X-ray crystal structure determination of OhrR and SlyA proteins and will be discussed further in next sections.

2. 3 *De novo* phasing methods: Multiple Anomalous Diffraction (MAD)

Since the first protein structure, basic blue copper protein, was solved by MAD in 1998 (Guss et al., 1988), the application of this method and the success of structure determination are rapidly growing. Even though many factors contribute to its success, the key to the success of MAD relies on the ability to select specific energies from the broad band of energies in the incident X-ray beam provided by a synchrotron source (Ogata, 1998). The theory of MAD phasing in crystallography and its practical aspects will be discussed in this and next following sections.

For MIR, heavy atom derivatives are often non-isomorphous with each other, which degrades the accuracy of the phases and even leads to the failure of this method. However in MAD experiments, the anomalous and dispersive differences, which are used to calculate the phases, are taken from the same crystal. In order to obtain anomalous signals such a crystal must contain an anomalously scattering atoms for determining phases. Selenium is anomalously scattering atom which is often used by incorporating seleno-methionine into a protein.

When the incident photon has a relatively low energy, the photon is either scattered or not, but is not absorbed as insufficient energy to excite any of the available electronic transitions. Thus, the photon scatters with no phase delay (imaginary, or f'', component is 0 in equation 2. 9). The scattering cross-section of the atom (or the probability that the photon is scattered) may be adequately described by using the normal atomic scattering coefficient f_o only. However, when the incident photon has high enough energy, it can promote an electronic transition between orbitals (discrete electronic energy levels). Some photons are absorbed and re-emitted at lower energy (florescence) or immediately re-emitted at the same energy (strong coupling to absorption edge energy). Therefore, three components need to be considered in anomalous scattering. First is normal (f_o), second, absorption (f') and third, anomalous (f'') scatterings which can be scattering reflected in the structure factor ($F_H(hkl)$) by following relationship:

$$f_o + f' + f'' = F_H(hkl) \qquad (2.\ 9)$$

For the f'' component, its phase is delayed by 90° (degrees) since the scattered photon gains an imaginary component to its phases (f'' scattering coefficient becomes non-zero).

For example, it is retarded compared to a normally scattered photon. This effect is most easily measured as a function of X-ray energy by noting either a sharp increase in absorption or in fluorescence at certain wavelengths. The imaginary scattering component f'' is proportional to these directly measurable quantities. The real scattering component f' is related to f'' via the Kramers-Kronig relationship;

$$f'(\omega') = \frac{2}{\pi} \int_0^\infty \frac{\omega' f''(\omega') \, \partial\omega'}{\omega^2 - \omega'^2} \qquad (2.\ 10)$$

Once the f″spectrum is obtained experimentally from the sample crystal via X-ray fluorescence measurements, the corresponding f′ spectrum can by calculated by numerical integration using the above equation (Figure 2. 3) (Hendrickson et al., 1988).

For normal X-ray diffraction, Bragg reflections related by inversion through the origin (*hkl* and *–h-k-l*, also called as "**Friedel pairs**") have equal amplitude and opposite phase. This phenomenon has been described as "Friedel's Law";

$$|F_F(hkl)| = |F_F(-h-k-l)| \qquad \alpha_H(hkl) = -\alpha_H(-h-k-l) \qquad (2.\ 11)$$

where F is the structure factor and α is the phase. However, when there is an anomalously scattering atom, Friedel's Law is broken (Figure 2. 5). This is because the f″ term is always positive.

$$|F_H(hkl)| = |F_H(-h-k-l)| \qquad \alpha_H(hkl) \neq -\alpha_H(-h-k-l) \qquad (2.\ 12)$$

In practice we measure F_{PH} when crystals contain anomalously scattering atoms.

$$F_P(hkl) + F_H(hkl) = F_{PH}(hkl) \qquad |F_{PH}(hkl)| \neq |F_{PF}(-h-k-l)| \qquad (2.\ 13)$$

And the observable anomalous difference, ΔF_H, is

$$\Delta F_H = \big|\, |F_H(hkl)| - |F_H(-h-k-l)| \,\big| \qquad (2.\ 14)$$

For the same reason as the members of a Friedel pair, Bragg reflections which are space group symmetry equivalents to the two members of a Friedel pair, called as "Bijvoet pairs", have unequal amplitudes in the presence of anomalous scattering. The difference in the measured amplitude for a Bijvoet pair is called a Bijvoet difference (Drenth, 1999).

A **Patterson map** that is calculated from the Bijvoet difference (with coefficients), ΔF_H, contains peaks corresponding to interatomic vectors between pairs of anomalous scattering atoms. The Patterson function is given by:

$$P(u, v, w) = 1/V \, \Sigma_h \Sigma_k \Sigma_l \, \Delta F_H^2 \exp\{-2\pi i (hu + kv + lw)\} \qquad (2.15)$$

where u, v and w are fractional coordinates in Patterson space. This equation gives the relative positions of the anomalous scattering atoms by using only the measurable square of the structure factor, not phases. This is the first step in MAD phasing, since the location of the anomalous scattering atom is needed to develop MAD phase estimates.

Knowing the phase angle calculated from the anomalously scattering atom, we can get the length and phase of F_H(hkl) and F_H(-h-k-l). Measuring the intensity of different data sets, we can get the length of F_{PH}(hkl) and F_{PH}(-h-k-l) and treating one wavelength data set as a native or reference provides the length of F_P (hkl). By constructing a conventional phasing diagram where the head of F_H (hkl) vector is at the origin, intersecting points of a circle with a diameter, $|F_P(hkl)|$, represent two possible choices of corresponding phase angle, αF_P (hkl).

$$F_P(hkl) = F_{PH}(hkl) - F_H(hkl) \qquad (2.16)$$

By Fridel's law

$$F_P(hkl) = F_P(-h-k-l) = F_{PH}(-h-k-l) - F_H(-h-k-l) \quad (2.17)$$

By applying the information of $F_H(-h-k-l)$ the phase can be determined (Figure 2. 6). This is called as the "**Harker construction**" (Drenth, 1999; Schumacher, 2004; Stout and Jensen, 1989).

The changes in the intensities caused by the anomalous scattering are very tiny. Differences of only a few percent can be observed between $|F_H(hkl)|$ and its Friedel pair $|F_H(-h-k-l)|$ at the same wavelength (Bijvoet differences) and between the same $|F_H(hkl)|$ at different wavelengths (called as "**Dispersive difference**") (Terwilliger, 1994). The question then arises how do we estimate these anomalous effects and how do we accurately need to measure the data to distinguish between an actual dispersive signal and just noise in the data. The success of MAD phasing critically depends on data quality (high signal to noise ratio) and completeness, and therefore special data collection techniques for maximizing the Bijvoet differences from Friedel pair (inverse beam) etc. are advantageous.

Once the phases have been determined an electron density map can be calculated. However errors in phase determination result in errors in electron density maps. In order to minimize the errors in phase determination a weighting factor, termed the "**figure of merit**" is applied to find the average structure factor, termed F_{hkl} (best) by:

$F_{hkl}(\text{best}) = m \mid F_{hkl} \mid \exp(i\alpha_{best})$

where m is a figure of merit and α_{best} is the phase from the average structure factor (Figure 2. 7). Thus, reduction in structure factor amplitude is expressed through the

figure of merit. For perfectly defined phases, the figure of merit is 1. As the phase information becomes more ambiguous, it drops, until it becomes zero when all phases are equally probable.

2. 4 In Practice: *de novo* structure determination by MAD

In practice *de novo* macromolecular X-ray crystal structure determination can be done through a series of multiple steps summarized in Figure 2. 8. After preparation of a pure, functional and homogeneous protein of interest, crystallization trials of the protein are performed by the hanging drop-vapor diffusion technique, one of most common methods of crystallization (Figure 2. 8A). In this method, a mixture of a protein and a crystallization solution (usually 1:1 ratio) is placed and suspended on a glass coverslip over a larger reservoir of the crystallization solution. Then water diffuses out of the protein drop and into the reservoir, gradually concentrating the protein and the precipitant in the drop and finally crystals can be formed. The *de novo* crystallization experiments to obtain diffraction quality crystals employ an incomplete factorial or sparse matrix approach. Initial crystallization screens will test the effects of wide ranges of pH, ionic strength, counter ions, temperature and a variety of precipitating agents, which include PEG, ammonium sulfate, sodium/potassium phosphate and 2-methyl-2, 4-pentanediol(McPherson, 1999). Once crystals are grown, the diffraction quality of the crystals is screened by X-ray diffraction (Figure 2. 8B) either in house or at an available synchrotron source, i.e., at Stanford Synchrotron Radiation Laboratory (SSRL) or Advanced light source (ALS) at Berkely, CA and X-ray intensity data is collected.

For data collection most crystals are frozen. Freezing crystals can reduce X-ray damage, particularly at synchrotron sources where the flux is very high. MAD data collection, which requires that multiple data sets be collected from a single crystal, must be done with a frozen crystal. Crystals can be frozen in many different ways. The easiest and most common method is to freeze the crystal in a liquid nitrogen cold stream (temperature is 100K) using cryoprotectants, which are necessary to prevent ice formation in the crystal. Since macromolecular crystals are made up of macromolecules as well as solvent molecules. Most protein crystals contain between 30 to 70 % solvent contents. Upon freezing, the formations of ice where solvent molecules are present, is very detrimental to both data collection and the crystal lattice. When ice crystals grow within a lattice, the crystal becomes disordered because ice is less dense than liquid water and ice diffracts X-rays as well. Typically, glycerol, ethylene glycerol, PEGs, MPD and oils are added as cryoprotectants but obtaining the best cryo-condition is a process of trial-and-error and truly empirical. Even in a frozen crystal, radiation damages inevitably occur after a long period of time for data collection (Weik et al., 2000).

After collecting initial X-ray data, both intrinsic symmetry (space group) and crystal orientation in the X-ray beam need to be determined. This will determine the optimal amount of data to collect before the crystal gets damaged by radiation and without wasting time and effort. These days, the determination of the space group and crystal orientation beam can be aided by many available computer programs (*e. g.*, Mosflm (Leslie, 1999)) using an auto-indexing process. Indexing means the assignment of a consistent set of three reciprocal space vectors, which define the reciprocal lattice

represented by the diffraction spots. The corresponding real space vectors (a, b and c) and the angles between them (α, β and γ) define the crystal unit cell. Next, the entire raw data set needs to be further merged, sorted and reduced into a unique set based on the space group symmetry and can be done using a computer program, Scala (Evans, 1997) as implemented in CCP4 (CCP4, 1994). If multiple data sets are used for phasing, such as in MAD phasing, these data sets must be brought onto a common scale as well.

At this stage the possible number of molecular subunits in the asymmetric unit of the crystal can be estimated. An estimation of the number of molecules per unit cell (Z) can be made by a method proposed by (Matthews, 1968). He found that for most protein crystals the ratio of the unit cell volume and the molecular weight is between 1.7 and 3.5 $Å^3$ / Da with most values around 2.15 $Å^3$ / Da. This number is called the Matthews' coefficient or V_M. This number is a good indicator of how many molecules reside within the asymmetric unit.

$$V_{protein} = (\text{Volume of protein in the unit cell})/V_{cell}$$

$$= \frac{(Z \times Mr \times \text{specific volume of the protein})/N}{V_M \times Z \times M_r}$$

$$= \text{specific volume (in } cm^3/g) \,/\, V_M \text{ (in } Å^3 / \text{Da} \times N \text{ mol}^{-1}) \qquad (2.\ 18)$$

where unit cell volume V_{cell}, Avogdro's number N and molecular weight M_r. The specific volume of a protein molecules is always approximately 0.74 cm^3 /g, and gives:

$$V_{protein} = 1.23 / V_M \qquad V_{solvent} = 1 - 1.23 / V_M \qquad (2.\ 19)$$

Thus from the V_M, the solvent fraction in the unit cell can be estimated (Drenth, 1999). In case of protein-DNA complex crystals, which tend to deviate from the typical range of

solvent contents, the answer may not be unambiguous and must await the phase information.

In order to determine the phase (Figure 2. 8C) by MAD experiment, either selenomethionine-substituted protein crystals or 5-bromouracil incorporated DNA in a nucleic acid or protein-nucleic acid complex crystals are most common ways to prepare crystals containing anomalous scattering atoms. In selenomethionine-substituted crystals, selenometionine is incorporated in place of methionine residues. For preparation of selenomethionine substituted protein, the protein is expressed utilizing the methionine inhibitory pathway. Analysis of the success of the substitution of methionine for selenomethionine can be monitored indirectly by amino acid analysis because selenomethionine is destroyed under conditions used for amino acid analysis (Doublie, 1997). Or when a selenomethionine substituted protein crystal is available the presence of a selenium peak in an X-ray fluorescence scan at the selenium edge can be used.

As mentioned earlier the magnitude of the differences at the structure factors corresponding to anomalously scattering atoms in the unit cell is very small. Thus the wavelengths to collect MAD data sets must be selected very carefully. Once the actual scattering factors f' and f'' in the energy range of interest for your sample crystals are determined, the exact x-ray wavelengths to use for MAD data collection can be chosen. For a single type of anomalous scattering atom, i. e., a MAD experiment at a single absorption edge, a minimum of two wavelengths is required. However, it is better to have more data points so that the set of simultaneous equations for MAD phasing is over-

determined, so three wavelengths are typically collected. The largest anomalous signal will come from choosing the wavelength with maximal f'' (λ_1, peak). The second wavelength is usually chosen to have maximal $|f'|$ (λ_2, inflection) to achieve the largest dispersive differences between f' and f'' (Figure 2. 4). Because λ_1 and λ_2 are very close together, a possible great precision is required in setting up the apparatus which controls wavelength during data collection. An additional wavelength (λ_3, remote) is chosen at a point remote from the absorption edge to provide a low anomalous difference signal but with high dispersive contributions. Typically λ_3 is between 100eV and 1000eV from the absorption edge. Considering the effects of radiation damage during data collection, which causes a loss of signal and crystal diffraction, the most common strategy is to obtain peak wavelength data first, followed by inflection point data and finally a remote data set (with redundant internal anomalous differences and large dispersive differences against inflection data). To obtain the data sets with complete Friedel pair differences, at the same time increasing the data redundancy, an inverse beam strategy is commonly used as well. This involves collecting equivalent data 180° away from the initial orientation.

Once MAD data sets are collected, these data are used to calculate difference Patterson maps. However, Patterson peaks are in *u*, *v* and *w* coordinates, not real-space coordinates and are often noisy because the number of Patterson peaks are equal n^2-n, where n is the number of atoms in the unit cell. Therefore, the use of "Harker sections" is necessary. Harker sections are created due to symmetry relationships. For an example, in the $P4_12_12$ space group, to which a reduced OhrR crystal belongs, an atom at *x*, *y* and *z* is equivalent

to a symmetry related atom at $-x$, $-y$, and $z + 1/2$. The relationship between u, v and w and x, y and z, follows:

$$u = -x - x = -2x \qquad x = -u/2$$

$$v = -y - y = -2y \qquad y = -v/2$$

$$w = z + 1/2 - z = 1/2 \qquad w = 1/2 \qquad (2.\ 20)$$

Thus the corresponding Patterson peaks on the Harker sections, at positions of u, v and w, can be converted in real-space, (x, y and z) coordinates. Once the anomalously scattering atom sites are found, the phases can be calculated. However, it tends to be more complicated to find real Patterson peaks when there are many sites to be found in the asymmetric units or in a high symmetry space group.

The search for possible Patterson peaks (anomalously scattering atom sites) from anomalous Patterson maps for correct phase calculation can be aided by an automated computer program called SOLVE (Terwilliger and Berendzen, 1999). First, SOLVE calculates difference Patterson maps and find peaks. Then using a difference Fourier analysis, SOLVE identifies further anomalous scattering atoms. The difference Fourier function is real-space (x, y and z) coordinates:

$$\Delta\rho\ (x, y, z) = 1/V\ \Sigma_h\Sigma_k\Sigma_l\ \left|F_H(hkl) - F_p\ (hkl)\right|^2 \exp\{-2\pi i\ (hu + kv + lw) + i\alpha_{hkl}\} \qquad (2.\ 21)$$

where α_{hkl} are the phases calculated from the partial selenium sites. SOLVE evaluates those peaks by using Z-scores. After omitting some sites z-scores are calculated on the basis of agreement with the Patterson, difference Fourier, the figure of merit of phasing and defined solvent and macromolecules in a native electron density map (native Fourier map). By choosing the highest peak in each step SOLVE writes out final phases for top

solutions. The advantages of SOLVE scoring and phasing are that potential heavy-atom solutions (selenium sites) can be examined both with respect to their agreement with the Patterson function and with respect to the qualities of the resulting native Fourier, difference Fourier and figure of merit. In addition, the examination of the native Fourier not only yields information on the overall quality of a solution but also can often positively identify the hand of the heavy-atom solution. The correct map with the correct hand has features expected of a protein: regions which are flat (solvent) and other regions which have high variation (the protein) (Terwilliger and Berendzen, 1999).

Once the initial electron density maps are calculated using the anomalous scattering atom phases, the model has to be built into the electron density (Figure 2. 8D). Unfortunately, the initial electron density map tends to be hard to interpret in most cases. Through a process of density modification the interpretability of maps can be improved. One of the most powerful tools to obtain readily interpretable maps to the crystallographer is solvent flattening (or solvent flipping) (Cowtan, 1998), which can greatly improve the map. As mentioned earlier, the presence of bulk (non-ordered) solvent between molecules (about 30 to 70 % of the entire crystal volume) introduces a uniform density distribution in that region. Setting the solvent electron density to a constant value ("flattening"), in repeated cycles with adjustments in the solvent mask leads to drastically reduced phase angle errors and hence, clearer and more easily interpretated electron density maps (Terwilliger, 1999).

For visualizing models and maps and for building models into maps, the program O (Jones et al., 1991a) is used. A common strategy is to identify the main chain first, producing an alpha carbon trace. Regions of secondary structures, alpha helices and beta sheet, can often be identified at this stage. If the resolution is sufficient, side chains can be placed. The locations of bulky side chains and the anomalous scattering atom sites, i. e., selenium sites, are very useful determine the register of the protein side chains (Figure 2. 8C and D). Typically the map improves throughout the refinement, which is the process of adjusting the model to agree better with the measured data. Refinement is done by improvement of the **R-factor** ("residual-factor" or agreement factor). The R-factor tells the average fractional disagreement between model and measured data and is defined by:

$$\text{R-factor} = \frac{\sum_{(hkl)} \left| \left| F_{obs}(hkl) \right| - \left| F_{calc}(hkl) \right| \right|}{\sum_{(hkl)} \left| F_{obs}(hkl) \right|} \qquad (2.\ 22)$$

where $F_{obs}(hkl)$ is observed structure factor and $F_{calc}(hkl)$ is calculated structure factor. When the electron density maps are calculated at moderate resolutions, it is very possible to overfit the data or to introduce model-bias. This problem can be avoided if a randomly chosen partial data set (test set, usually 5 ~ 10 %) set is set aside along from the refinement of most of the data (working set) to refine the atomic model and reduce model bias. The test set data are used to compute R_{free} (for working set, R_{work}), which is computed in the same way as the conventional R-factor but using only that subset of data. If R_{free} drops at the same time when R_{work} does then the model must really have improved. This idea, called as cross-validation is employed in refinement as implemented in the crystallography&NMR system software (CNS) (Brünger et al., 1998).

During model building and refinement procedures, typically used electron density maps are $2F_o$-F_c and $F_o - F_c$, where F_o is observed structure factor and F_c is calculated structure factor. By substracting F_c from $2F_o$, $2F_o$-F_c maps exaggerate the areas where F_o differs from F_c, while still yielding a map that resembles a molecular surface. In $F_o - F_c$ maps (or called as difference maps) "positive" or "negative" peaks are produced in area where F_c differs from F_o. Such maps are very useful to determine any missing part at positive contouring and wrong part at negative contouring. This map is usually contoured at a high level between - 3 and - 4 σ to view the large difference peaks against noise (Tao, 2005). Simulated annealing omit maps (Brünger et al., 1998), where a randomly chosen part of the model is omitted then the map is is generated for remaining part (typically $2F_o$-F_c map), are used to reduce model bias.

A **B-factor**, which is the "temperature-factor" or "Debye-Waller factor", is also applied to determine the thermal motion of the atoms. The B-factor, which is applied to the X-ray scattering term for each atom (or groups of atoms, depending on the highest resolution) describes the degree to which the electron density is spread out. While the theory is that the B-factor indicates the static or dynamic mobility of an atom, it can also indicate where there are errors in model building. The B-factor is given by:

$$B_i = 8\pi^2 U_i^2 \qquad (2.\ 23)$$

where U_i is the mean square displacement of atom i in Å. This produces a weighting factor (weakening effect) on the contribution of atom i to the Fourier transform by;

$$f_{Bi} = f \bullet \exp(-B_i \frac{\sin 2\theta}{\lambda^2}) \qquad (2.\ 24)$$

As U increases, B increases and the contribution of the atom to the scattering is decreased. If atoms are incorrectly built, their B-factors will tend to be much higher than correctly built atoms nearby.

After model building and refinement, the macromolecular structure needs to be validated in order to examine the model quality (Figure 2. 8E). Model-quality is analyzed by several criteria based on the use of the *x*, *y*, and *z* coordinates of the atoms and their properties (*e. g.*, carbolyxate oxygen or aromatic carbon), and their identities (*e. g.*, the main-chain nitrogen atom of a certain residue). Often these criteria compare properties of the model against the standard sets of bond length and bond angles libraries between atoms, which are based on chemistry, physics and a large collection of macromolecular structure databases (Kleywegt, 2001). Certainly the R-factors and B-factor are both great indicators used to describe how the structural model fits the experimental data. In addition, to check the stereochemistry of a protein model the "Ramachandran plot" is used. For a polypeptide the main chain nitrogen-alpha carbon and alpha carbon-carbon bonds free to rotate in certain ways. These rotations are represented by the torsion angles phi and psi, respectively. The ramachandran plot shows the analysis of phi and psi angles of a polypeptide model, in which only certain regions are in allowed stable conformations. Close contacts between atoms (spheres corresponding to their van der Waals radii are used in Ramachandran analysis) are not sterically allowed. However, in order to form a stable protein in a globular structure, sometimes, residues (mostly located

at a loop or turn) with unusual phi and psi angles are found. Another program, PROCHECK (Laskowski et al., 1993) is a nice program for evaluating models in many ways including planarity of peptide bonds, root mean standard deviation of the atoms' bond lengths and angles against a standard library, chirality and Ramachandran analysis.

2. 5 Molecular Replacement (MR)

MR is generally the most rapid method in solving structures of the same or similar molecules in a different crystal form (different space group). In this method molecular coordinates (M) predicted as similar or identical to the unknown structures (X) are required and such structure (M) is called as a search model or template. Because of the rapidly growing number of protein structures that are available as search model, MR method has been successful in recent years. The template structure is chosen by primary sequence similarities because protein sequences can tell us whether the protein structures are likely to be the same or not. If the sequence similarity is over 30 %, then we say the sequences are homologous, meaning they likely evolved from the same common ancestor, thus they may have similar structure. However the primary sequence similarity alone dose not guarantee that an unknown structure can be solved by MR. Any subtle structural dissimilarity can perturb the structure enough for MR to fail.

The initial phases are calculated based on the positions of all the atoms in the MR template (M) and such phases are often sufficient to obtain a usable electron density map that can be used to refine the structure of the target protein. To obtain phases of X requires decisions about the boundaries that identify a molecule. There is a very simple

case where the crystal symmetry is unchanged and the unit cells are virtually the same. Such cases may be studied by electron-density difference maps, similar to those described for the assignment of heavy atom positions in MIR method. This is quite common in the case of single-site amino-acid substitutions in a protein and a ligand or a drug bound in a protein crystal and the difference map may show the electron-density differences at the substituted amino acid and at the ligand binding site, respectively, while the absence of the significant density in the rest of the map could confirm that negligible changes had occurred in the rest of the structure (Blow, 2002a).

In other case, the crystal forms of M and X crystallize differently, and the space-group symmetry is probably different. Or the crystal forms M and X belong to same space group but they are non-isomorphous. Such case requires new phases. It is important to begin by determining the symmetry and cell dimensions of the X crystals. With this information, we can estimate the number of M-like molecules in the asymmetric unit of X crystals by calculating Matthews coefficient.

The next step to solve structure by MR is a rigid body transformation of M in crystal X. The word "rigid" means that the movement has not distorted it in any way. Conventionally this displacement can be divided two steps. First, rotate the original molecule about the reference point until it is exactly parallel to the displaced molecule. Then, move it without any rotation to the displaced position. The position of molecule in the crystal unit cell can be represented using 6 parameters, 3 angles of rotation (defining a

matrix of 9 coefficient) and a vector of translation (3 values), resulting in 6-dimensional space search of its current position, *i.e.*

$$x' = c_{11}x + c_{21}y + c_{31}z + v_x$$
$$y' = c_{12}x + c_{22}y + c_{32}z + v_y$$
$$z' = c_{13}x + c_{23}y + c_{33}z + v_z$$

However, to search for the correct displacement, a six-dimensional search is required which is a demanding and expensive problem. For example, even if only 10 values were considered for each variable this would generate a 10^6 search points to be examined. Fortunately, application of the Patterson function allow the searches for rotation and translation to be done separately, thereby breaking the 6-dimensional search into two three-dimensional ones (Rowlett, 2005).

The rotation function is based on a **self-Patterson** which involves the intramolecular 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. The rotation function is defined as the overlap of one self-Patterson function. If a number of identical molecules (or subunits) lies within one asymmetric unit, the self Patterson vector distribution is exactly same for all these molecules, expect for a rotation that is the same as their noncrystallographic rotational symmetry in real space, Therefore, if the patterson function is superimposed on a correctly rotated version, maximum overlap between the tow Patterson maps will occur. Similarly, for two different lattices the two different Patterson maps must be superposed to maximum overlap by a rotation of one of the two maps (Drenth, 1999).

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. Rossman and Blow applies first a rotation around z-axis by angle α, then progresses around the new position of x-axis by angle β then around z-axis by angle γ, called as zyz convention. Molrep (Vagin and Teplyakov, 1997), the program used in this thesis uses this function to increase the speed of rotational sampling.

From the result of a rotation study the model molecule M is rotated so that is lies parallel to one of the molecules in crystal X. Next, the translation search will proceed to determine the proper position of structure (three positional parameters) in the unit cell only exists if there is some crystal symmetry. In P_1, the triclinic crystal system there is no symmetry and the origin can be chosen to be anywhere. But if there is rotational symmetry in crystal X, the crystal origin must be defined related to the position of its rotation axes.

The translation function is basically a trial and error process involving the rotated molecule through asymmetric unit in crystals and calculating the structure factor amplitudes of the search model (F_{calc}) in every new position and compared to that of the observed data (F_{obs}). An alternative to the trial and error method in the translation function is based on a **cross-Patterson** vectors (intermolecular vectors) that relate atoms in two different molecules related by crystallographic symmetry so that phase information is not needed. In essence, the search consists in placing the origin of

molecule M in the unit cell of X. For each position the predicted configuration of Patterson vectors is compared to the actual Patterson function of X. Compared to the rotation function, a translation function can be computed very rapidly. When the correct translation is found there should be a large peak because the vector sets form the model and the data will coincide.

When the rotation and translation operations are known, these can be applied to all atoms of the model M, placing them at appropriate positions in the unit cell. Estimated structure factor of the crystal structure can then be calculated. R-factor or correlation coefficient between these F_{calc} and F_{obs} is a good indicator to find best solutions in the translation function.

$$R = \frac{\Sigma \left| \left| F_{obs} \right| - \kappa \left| F_{calc} \right| \right|}{\Sigma \left| F_{obs} \right|}$$

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

In most cases, R-factor wants to be a low number, lower than 0.5. However, the correlation coefficient needs to close to 1.0 meaning that the correlation between any two functions is perfect.

2. 4 In Practice: structure determination by Molecular Replacement (MR)

After collecting a native data set, it is desirable to select a MR protein that is as homologous as possible to the target protein and to examine sequence alignment of the two proteins. MR method may be possible if the proteins are more than 30 % identical. Usually the MR template protein needs to be modified to make it a similar as possible to the target protein. As an example, if the MR protein has extra residues, either N- or C-termini, any extra part needs to be removed. Or only homologous domain can be used for MR template. If there are significant mismatches in primary sequence alignments or X-ray diffraction data have mild resolution, lower than 3 Å, sequences of MR protein need to be mutated to Alanine to neglect the error introduced by the side chain in search process.

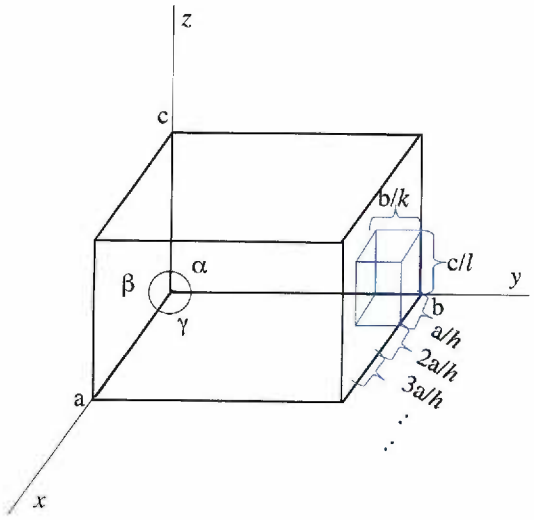
Evolutionary programming for molecular replacement (EPMR) (Kissinger et al., 1999) uses an efficient 6-dimensional evolutionary search algorithm (similar to genetic algorithm) to one of many good fits of the search model to the reflection data during each trial. The search is repeated for many trials, starting with different initial orientations of the search model. The results of the best of these trials is assumed to be close to global best fit, providing a good model for estimating phase data and constructing the first electron density map.

A decent molecular replacement solution will have an R-factor no larger than 0.45. If R > 0.50 it is unlikely that the MR solution sill be useful. If the R-factor is satisfactory, then the molecules placed in the unit cell should be examined for overlaps with themselves or with symmetry-generated partners by loading them into O (Jones et al.,

1991a). Of there are no obvious overlaps and the symmetry-generated molecules pack well into the unit cell, model refinement or further MR trial proceeds.

After initial rigid body refinement, the model is typically subjected to simulating annealing, which essentially "heating and shake up" the model in a random way, followed by "cooling" to find a better but less model-biased fit to the experimental data. After this step the model is typically taken through repeated rounds of whole-molecule minimization, b-factor refinement and the generation of new electron density maps for visualization and manual adjustments in order to make the model fit into better compliance with electron density maps. Typically as the refinement proceeds, it may become obvious that some portions of the proteins are not visible, and should be removed. Alternatively as the refinement proceeds, new regions of electron density maps may become obvious and clear allowing the addition of residues to the model, especially at the N- and C- termini (Rowlett, 2005). All these refinement processes can be done in CNS (CCP4, 1994).

A



B

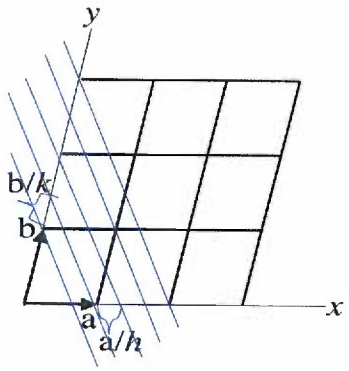


Figure 2. 1 **A**. A unit cell with six parameters in *x*, *y* and *z* and α, β and γ. A unit cell is shown as a black box. The blue box indicates the unit cell is cut in a/*h* pieces in the *x*-axis, in b/*k* in *y*-axis and c/*l*, in *z*-axis. **B**. Two dimensional view of a lattice in *xy*-plane. Real and reciprocal lattice planes are indicated in black and blue lines, respectively. The reciprocal lattices cut the ab-face of the unit cells by one third in *x*-axis and half in *y*-axis generating the 32 reciprocal lattices. Adapted from (Drenth, 1999)

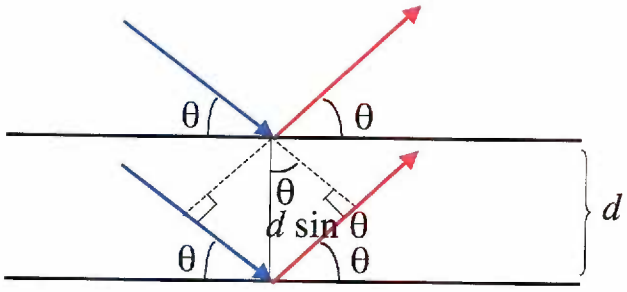


Figure 2. 2 Bragg's law. X-ray diffraction can be viewed as a process that is similar to reflection from plane of atoms in the crystal. The incident and scattered X-rays (wavelength, λ), shown as blue and red arrows, respectively from the same plane in black horizontal lines, with d spacing normal to the crystal planes. The path length will be $2d$ sin θ. Thus for constructive scattering (diffraction) to occur, Bragg's law ($n\lambda = 2d \sin \theta$) must be satisfied. For a given d spacing and wavelength, this first order peak ($n = 1$) occurs at a particular θ value. Similarly, the θ values for the second ($n = 2$) and higher order ($n > 2$) peaks can be predicted. From (Powell, 2005)

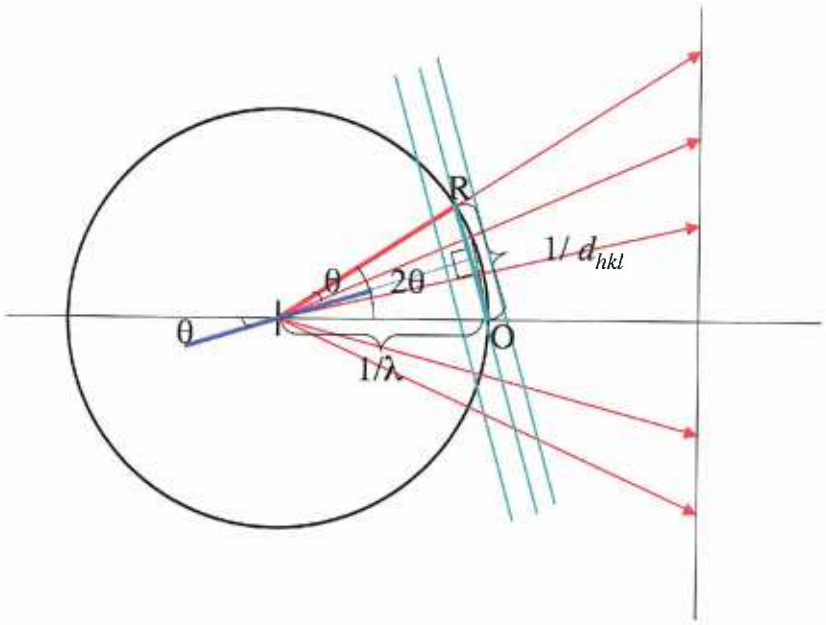


Figure 2. 3 Two dimensional representation of the generation of an X-ray diffraction patterns and Ewald's sphere. The circle is Ewald's sphere with a radius of 1/λ. The planes, shown in green lines, are part of a lattice composed of diffraction spots. O indicates the origin of the reciprocal lattice and line OR represents the reciprocal lattice line with *hkl* in reflection position. Note that diffraction occurs only if the spots are on, or pass through, the surface of the sphere. Adapted from (Drenth, 1999)

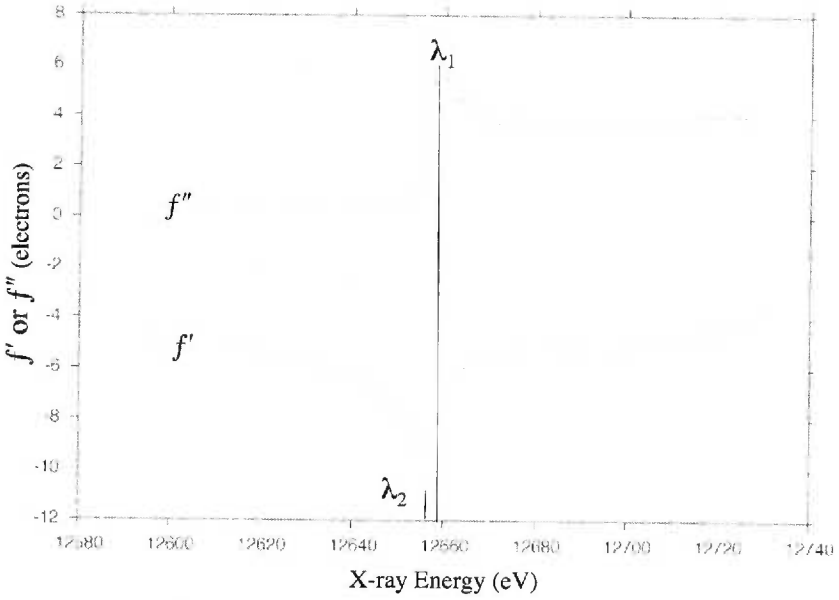


Figure 2. 4 X-ray selenium fluorescence scan of reduced OhrR protein crystal containing selenometionine-substituted OhrR protein. Wavelengths are chosen at the peak of the f'' (λ_1 = 12658.60 eV), a maximal $|f'|$ (λ_2 = 12656.08 eV) and the third wavelength (λ_3) is chosen at 13500 eV for actual data collections.

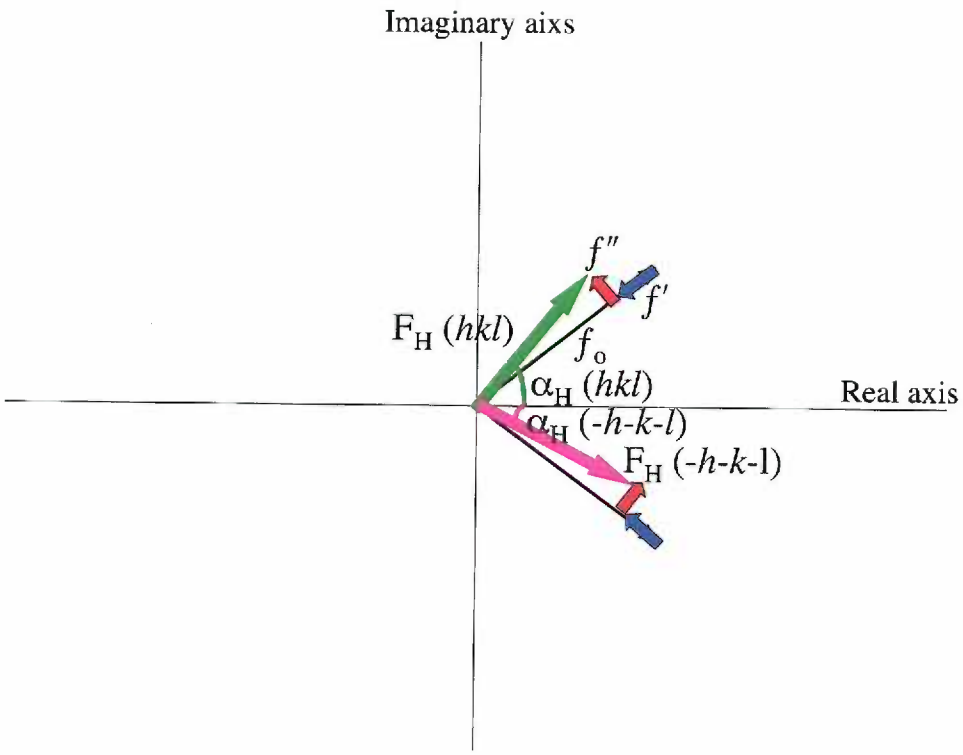


Figure 2. 5 Anomalous scattering. $F_F(hkl)$ and $F_F(-h-k-l)$ are Fridel pair. f_o is denoted by black arrow; f', blue arrow; f'' red arrow; $F_F(hkl)$, green arrow; $\alpha_F(hkl)$, green line; $F_F(-h-k-l)$, pink arrow; $\alpha_F(-h-k-l)$, pink line. From (Schumacher, 2004)

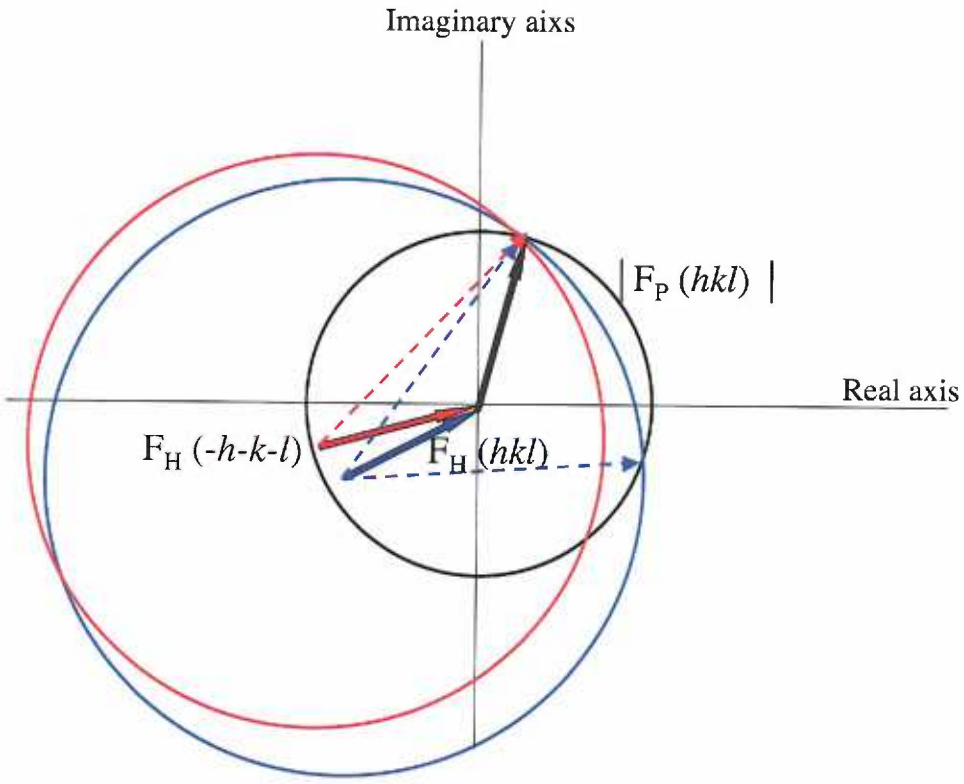


Figure 2. 6 Phase determination by the Harker construction. Black circle has a diameter of $|F_P\ (hkl)|$; blue circle, a diameter of $|F_{PH}\ (hkl)\ |$; red circle, $|F_{PH}\ (-h-k-l)\ |$. Phase ambiguity can be solved by this equation; $F_P\ (hkl) = F_{PH}\ (hkl) - F_H\ (hkl) = F_P\ (-h-k-l) = F_{PH}\ (-h-k-l) - F_H\ (-h-k-l)$. From (Schumacher, 2004)

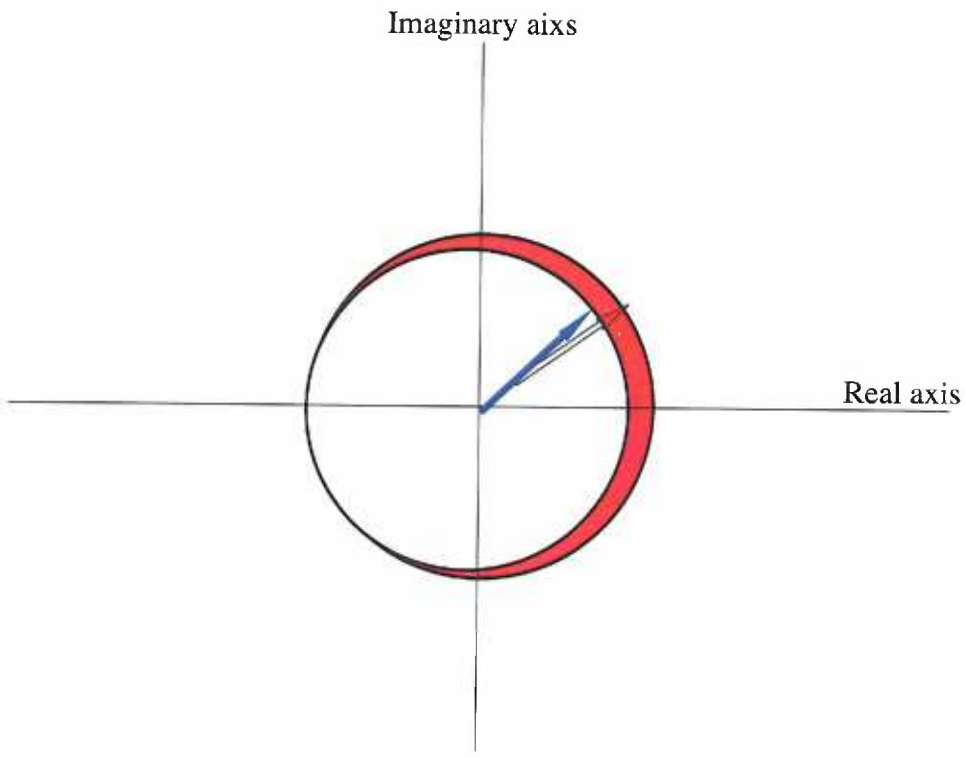


Figure 2. 7 Figure of merit. The circle represents possible values for the structure factor F_{hkl}. The probability of each possible phase is indicated by the thickness of the red line around the circle. Averaging a structure factor around a circle gives F_{hkl} (best) shown in a blue arrow inside the circle with a corresponding "best phase, α_{best}".

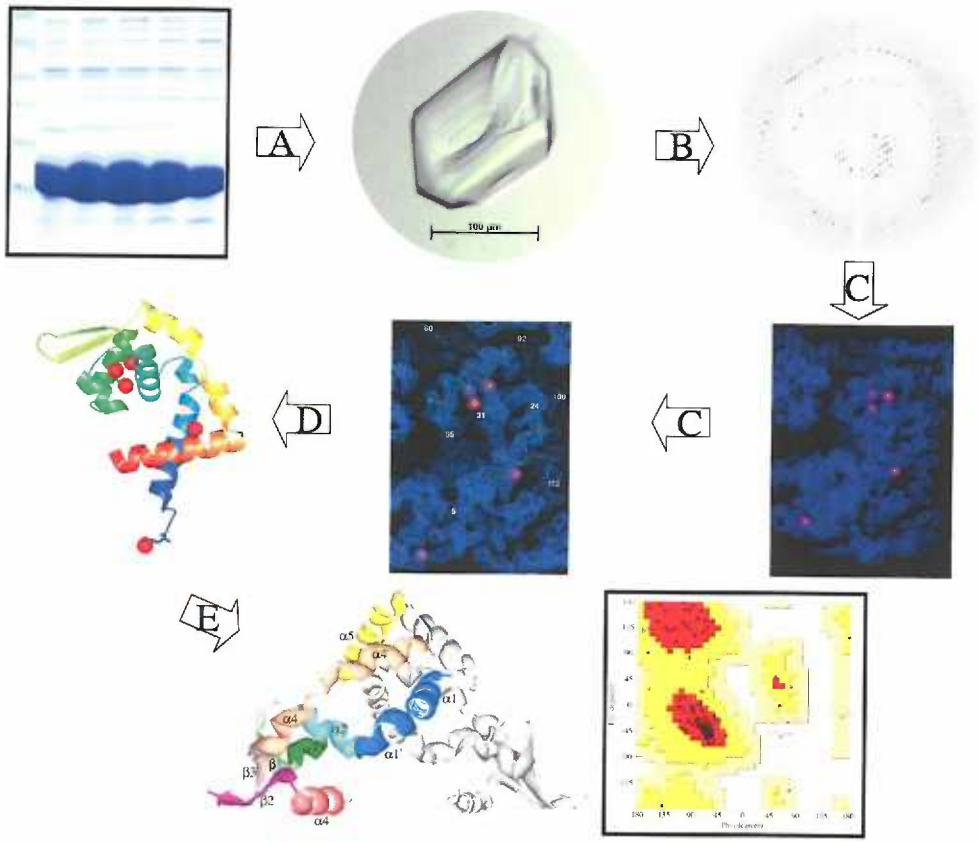


Figure 2. 8 A simplified flow-chart of the process of macromolecular structure determination by X-ray crystallography. All images are used from the *de novo* structure determination of reduced OhrR protein crystal. Arrows indicate steps described in text.

Chapter 3.

Manuscript #1

Structure of an OhrR-*ohrA* operator complex reveals the DNA binding mechanism of the MarR family

Minsun Hong[1], Mayuree Fuangthong[2,3], John D. Helmann[2] and Richard G. Brennan[1]*

[1]Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, Oregon 97239-3098, U.S.A.

[2]Department of Microbiology, Cornell University, Ithaca, NY 14853-8101, U.S.A

[3]Present address: Laboratory of Biotechnology, Chulabhorn Research Institute, Lak Si, Bangkok 10210, Thailand

*Corresponding author information

Department of Biochemistry and Molecular Biology, Oregon Health & Science University, 3181 S. W. Sam Jackson Park Road, Portland, Oregon 97239-3098, U.S.A.

Telephone number: 001-503-494-4427

Fax: 001-503-494-8393

E-mail: brennanr@ohsu.edu

Summary

The mechanisms by which *Bacillus subtilis* OhrR, a member of the MarR family of transcription regulators, binds the *ohrA* operator and is induced by oxidation of its lone cysteine residue by organic hydroperoxides to sulphenic acid is unknown. Here, we describe the crystal structures of reduced OhrR and an OhrR-*ohrA* operator complex. To bind DNA OhrR employs a chimeric winged helix-turn-helix DNA binding motif, which is composed of extended eukaryotic-like wings, prokaryotic helix-turn-helix motifs and newly described helix-helix elements. The reactivity of the peroxide-sensing cysteine is not modulated by proximal basic residues but largely by the positive dipole of helix $\alpha 1$. Induction originates from the alleviation of intersubunit steric clash between the sulphenic acid moieties of the oxidized sensor cysteines and nearby tyrosines and methionines. The structure of the OhrR-*ohrA* operator complex reveals the DNA binding mechanism of the entire MarR family and suggests a common inducer binding pocket.

Running Title: Structure of an OhrR-*ohrA* operator complex

Introduction

All living organisms are known to contain molecular systems that enable them to resist a variety of toxic substances and environmental stresses. Exposure to reactive oxygen species (ROS) and subtoxic levels of heavy metals invokes unique adaptive responses beyond those of general stress. Because ROS [including hydrogen peroxide (H_2O_2), superoxide anion, hydroxyl radical, and organic hydroperoxides (OHP)] can damage proteins, DNA, lipids, and membranes and potentially lead to mutations and cell death, they must be removed rapidly (Gutteridge and Halliwell, 1989; Storz and Imlay, 1999). ROS are generated as byproducts of normal aerobic metabolism and defense responses of both animals and plants (Babior, 1984). Although the detoxification of H_2O_2 and superoxide radical is well understood and their regulatory mechanisms reasonably well defined (reviewed in (Paget and Buttner, 2003; Storz and Imlay, 1999), those of organic hydroperoxides are less well known. These peroxides threaten cell integrity as they can attack long chain fatty acids and other nonpolar components of the cell membrane.

Upon exposure to ROS or disulfide stress agents, *Bacillus subtilis* induces an oxidative stress stimulon that is controlled by OhrR (Fuangthong et al., 2001), σ^B (Helmann et al., 2003), PerR (Herbig and Helmann, 2001) and Spx (Nakano et al., 2003). The organic hydroperoxide resistance protein regulator (OhrR), a member of the multiple antibiotic resistance regulatory (MarR) family, is a novel organic peroxide-sensing transcription factor that controls expression of the organic hydroperoxide resistance (*ohr*) gene, by binding to *ohrA* promoter elements (Figure 3. 1A) (Fuangthong et al., 2001). Ohr and OhrR are present in both Gram-positive and Gram-negative bacteria and are essential for

the response against lipophilic ROS by sensing the presence of OHP (Lesniak et al., 2002; Mongkolsuk et al., 1998; Ochsner et al., 2001; Sukchawalit et al., 2001). Recent structural and biochemical studies on Ohr have shown that this enzyme contains alkyl hydroperoxide reductase activity and detoxifies OHP by reducing these peroxides to alcohols (Cussiol et al., 2003; Lesniak et al., 2002).

Bacillus subtilis OhrR is a homodimer of 147 amino acid residues per subunit (MW ~17 kDa) and belongs to the MarR family of transcription regulators (Fuangthong et al., 2001). A unique and conserved cysteine residue, Cys15 (Figure 3. 1B), is oxidized to Cys-sulphenic acid (CysSOH) by OHPs, such as *tert*-butyl hydroperoxide and cumen hydroperoxide, which results in the induction of OhrR and derepression of *ohrA* (Fuangthong and Helmann, 2002). Because of its high reactivity, the CysSOH may take higher oxidation states such as Cys-sulphinic acid ($CysSO_2H$) and Cys-sulphonic acid ($CysSO_3H$). However, the formation of an intersubunit disulfide bond has not been observed in *B. subtilis* OhrR (Fuangthong and Helmann, 2002). By contrast, OxyR, the global regulator of the peroxide stress response in *Escherichia coli*, does form an internal disulfide bond between residues Cys199 and Cys208 that begins with the oxidation of Cys199 to sulphenic acid, resulting in the formation of intramolecular disulfide bond (Figure 3. 2) (Choi et al., 2001; Lee et al., 2004).

Members of the MarR family regulate the expression of resistance genes to multiple antibiotics, organic solvents, detergents and oxidative stress agents (Seoane and Levy, 1995; Sulavik et al., 1995). Despite low primary sequence similarity amongst MarR

family proteins, less than 25% on average, the members of this family show significant structural homology. X-ray crystal structures of several proteins of the MarR family, including MarR (Alekshun et al., 2001), MexR (Lim et al., 2002), SarR (Liu et al., 2001) and SlyA-like protein (Wu et al., 2003), have revealed a common triangular shape with winged helix-turn-helix (wHTH) DNA-binding motifs at two of the corners (Figure 3. 3A). However, superimposition of structures of several MarR family proteins highlights their conformational plasticity (Figures 3. 3B and C), as large differences are observed (Lim et al., 2002). Such plasticity likely facilitates the transitions between the DNA-bound, inducer-bound and apo states of each protein. However, the lack of appropriate structures, including that of a DNA-bound MarR family member, has hindered a full understanding of the DNA-binding, repression and induction mechanisms of the MarR family.

Here, we describe the crystal structures of reduced OhrR and an OhrR-*ohrA* operator complex. The structures reveal the chimeric nature of the wHTH motif of OhrR and a new two helix DNA binding element, both of which are proposed to be utilized by the entire MarR family to bind cognate DNA. Moreover, the structures reveal unambiguously the location of the oxidant sensing cysteine and the novel mechanism by which the chemical reactivity of the sulphydryl group is modulated. Finally, the initial steps of the induction mechanism can be readily envisioned.

Experimental Procedures

Protein preparation

Wild type and a mutant OhrR protein, in which the oxidant sensor residue C15, the lone cysteine of OhrR, is replaced by a serine, were overexpressed as described (Fuangthong and Helmann, 2002). This substitution allowed facile protein purification by overcoming the inherent rapid oxidation of C15. Serine 15-substituted OhrR is an excellent mimic for reduced wild type OhrR and shows DNA binding affinity identical to wild type but does not react with OHP. Protein purification was modified slightly from the published method by the addition of an anion exchange chromatography step prior to heparin sepharose column chromatography (Figure 3. 4). OhrR C15S, now termed reduced OhrR, was concentrated to 40-60 mg/mL in 25 mM Tris, pH 7.4, 50 mM NaCl and 1 mM EDTA and used immediately in crystallization experiments. Selenomethionine-substituted reduced OhrR was obtained utilizing the methionine inhibitory pathway and purified as above.

OhrR equilibrium binding to operator DNA

Fluorescence polarization experiments were done with a PanVera Beacon 2000 Fluorescence Polarization System (PanVera Corporation), utilizing the increase in polarization that is observed upon the binding of the operator DNA to the protein (Lundblad et al., 1996). Binding is assayed by titrating increasing amounts of OhrR protein into the operator DNA solution and measuring the millipolarization (mP) at each titration point. The data can is then fitted to the following equation to determine the binding constant (K_d):

$P=\{(P_{bound} - P_{free})[protein\}/(K_d)+\{protein])\}+P_{free}$

where P is the polarization measured at a given total concentration of OhrR; P_{free} is the initial polarization of the free operator; and P_{bound} is the maximum polarization of specifically bound to DNA. 5′-Fluoresceinated oligonucleotides corresponding to the *ohrA* operator site (*ohrA* wild type) and the *ohrA*-29 mer used in crystallization of OhrR-*ohrA* complex, were annealed in 10 mM sodium cacodylate, pH 6.5, by heating to 80 °C followed by flash cooling. Binding was assayed in a 1-ml volume at 27 °C with 0.2 nM fluoresceinated DNA and 1.0 μg/ml polu ([dI•C]) in 25mM Tris, pH 7.4, 50mM NaCl and 1 mM EDTA. For the OhrR wild type experiments, 5mM β-mercaptoethanol was added to prevent oxidation. After each addition of OhrR protein, samples were incubated in the Beacon instrument at 27 °C for 40 seconds before a measurement was taken. The 30 seconds of incubation allowed equilibrium to be reached. The millipolarization (mP) at each titration point represents the average of 4 measurements integrated over 5 seconds. Samples were excited at 490 mm and emission was measured at 530 nm. The purified OhrR wild type and C15S are fully functional and their DNA binding affinities are very similar with Kd of 5.1 ± 0, 5 nM and 2.9 ± 0.2 nM, respectively (Figures 3. 5 A and B).

Crystallization and data collection of reduced OhrR

Reduced OhrR was crystallized at room temperature using the hanging drop-vapour diffusion method and reservoir solutions of 30 to 33% PEG 1450, 0.1 M ammonium acetate and 0.1 M sodium acetate, pH 4.5 (Figure 3. 6A). Crystals grew fully in 2 to 3 days and took space group $P4_12_12$, with a = b = 55.4 Å, c = 77.2 Å. An isomorphous

crystal of reduced wild type OhrR was obtained in the same crystallization condition. Crystals were cryoprotected by a solution of 50% PEG 1450, 0.2 M ammonium acetate and 0.2 M sodium acetate, pH 4.5, and were flash frozen in a nitrogen cryostream. MAD and native intensity data were collected at the Advanced Light Source (ALS), beamline 8.2.1. All data were processed with MOSFLM (Leslie, 1999) and merged using SCALA as implemented in CCP4 (Table 3. 1). 10% of the data was set aside for cross validation.

Structure determination and refinement of reduced OhrR

The structure of reduced OhrR was solved by MAD methods (Hendrickson, 1991). Using intensity data from three wavelengths, five of the six selenium sites per OhrR monomer were located by SOLVE (Terwilliger and Berendzen, 1999). After solvent flipping (26.6% solvent content), as implemented in the crystallography&NMR system software (CNS) (Brünger et al., 1998), the map revealed clear density for α helices and bulky sides chains (Figure 3. 6B). After model building in O (Jones et al., 1991b), simulated annealing and rounds of positional refinement (Brünger et al., 1998) and model rebuilding resulted in convergence and an R_{free} of 28.3% at 3.0 Å resolution. Using EPMR (Kissinger et al., 1999), the MAD structure was used as the search model in molecular replacement trials to determine the structures of the reduced wild type OhrR and reduced C15S OhrR proteins, at 3.4 Å and 2.5 Å resolution, respectively. Refinement of the best solutions was carried out using CNS with one initial round of rigid body refinement followed by a round of simulated annealing. Subsequent rounds of positional and thermal parameter refinement were carried out for each followed by model rebuilding. The R_{free} of the wild type and reduced OhrR structures converged to 34.4%

and 29.5%, respectively. An overlay of these structures resulted in an rmsd of 0.2 Å, underscoring their structural identity, including the hydrogen bond network about the reactive residue C15 (Figure 3A). The model of the higher resolution reduced OhrR protein has excellent stereochemistry with residue E53, the only Ramachandran outlier (Nicholls et al., 1991). The reduced OhrR structure includes residues 8–144 and 36 water molecules.

Crystallization and data collection of the OhrR-*ohrA* operator complex

Multiple oligodeoxynucleotides, containing the *ohrA* operator site and ranging from 15 to 31 bps, were used in crystallization trials. Diffraction quality crystals were obtained with a pseudopalindromic 29 bp operator site that encompasses the *ohrA* operator and contains 5′-T overhangs (Figure 3. 7A). This crystallized sequence contained three nonconsequential substitutions to enhance the palindromic nature of the binding site (Figure 3. 7B). A fluorescence polarization based OhrR-*ohrA* operator binding assay revealed a K_d of 2.6 ± 0.3 nM (Figures 3. 5C), which is identical to that of wild type operator DNA (data not shown). The starting concentration of each OhrR C15S protein was 0.5 mM and mixed with DNA at molar ratios of 2:1 and 2:1.5 (protein monomer: dsDNA). Two crystal forms of the native, i.e., OhrR C15S, and selenomethionine-substituted OhrR-*ohrA* operator complex were grown from solutions of 0.1 M sodium/citrate phosphate, pH 3.7, 0.1 mM $MgCl_2$ and 5 to 8% PEG 8,000. Crystals grew in 6 to 8 hours but dissolved after 2 to 4 days. The crystals take the monoclinic space group, $P2_1$, with a = 81.9 Å, b = 81.0 Å, c = 109.3 Å and β = 100.2° or the trigonal space group, $P3_121$, with a = b = 80.4 Å and c = 182.0 Å. Both forms were cryoprotected by

gradually moving crystals into 0.3 M sodium/citrate phosphate, pH 3.7, 0.3 M $MgCl_2$ and 25% PEG 8,000. Both MAD and native x-ray intensity data were collected at ALS beamline 8.2.1 and processed with MOSFLM and merged using SCALA. Five percent of the structure factors were set aside for cross validation.

Structure determination and refinement

The structure of the $P3_121$ crystal form of the OhrR-DNA complex was determined by MAD methods. Using SOLVE, fourteen of the eighteen selenium atoms in the ASU were located. Solvent flipping (49% solvent content) revealed continuous electron density that allowed the fitting of one OhrR dimer and its bound 29 bp *ohrA* operator and an additional OhrR subunit. Peaks in difference fourier map between MAD and 5'-iododeoxyuracil containing data sets assist the register of *ohrA* operator sites (Figures 3. 8). After model building in O and positional refinement, the R_{free} converged to 33.6%. The structure of the $P2_1$ crystal from was then determined by molecular replacement using EPMR and the $P3_121$ MAD-derived OhrR-29bp complex structure as the search model. Molecular replacement revealed three independent OhrR dimer-29bp DNA complexes per ASU. After rigid body refinement, a round of simulated annealing, multiple rounds of positional and thermal parameter refinement and model rebuilding, the R_{free} converged to 27.7% and R_{work} to 21.5%. The $P2_1$ structure includes residues 9 - 145 of the six OhrR subunits, all 29 nucleotides of each strand of 3 double stranded DNA operators and 179 waters. The current model has excellent stereochemistry with no Ramachandran outliers (Table 1).

The determination of the pK_a of reactive C15: Site-direct fluorescence labelling by PDT-Bimane

The pK_a of C15 sulphydryl group was determined by its reaction with (2-pyridyl)-dithiobimane (PDT-Bimane), which releases pyridine-2-thione and absorbs at 343nm. A stock solution of PDT-Bimane was made in DMSO and the concentration was determined using absorbance extinction coefficient at 380 nm, ε_{380} = 5000 M $\cdot cm^{-1}$ in ethanol (Carlsson et al., 1978; Mansoor and Farrens, 2004). The reaction was initiated by the addition of 20 mM PDT-Bimane (1 μL) into a cuvette containing 1 mL of 12 μM of wildtype OhrR in 100 mM sodium phosphate/citrate. The tested pHs ranged from 4.5 to 8.5 in 0.5 increments. The formation of the pyridine-2-thione product was monitored at 343 nm at room temperature at 15 seconds intervals. As a control, the same experiments were performed with the OhrR C15S protein, which assessed the relative stability of the PDT-Bimane to spontaneous hydrolysis. The resulting curves showed insignificant breakdown at each pH. Absorbance data were fitted directly using the equation ($A_t = -(A_{max}-A_o)*(e^{-kobs*t})+ A_{max}$, where A_t represents the amount of absorbance at 343 nm at time t; A_{max}, the maximum absorbance; A_o, the absorbance of time t = 0; k_{obs}, the rate constant at any given pH) gives the pseudo first order rate constant of the modification reaction (k_{obs}) at each pH with the program KaleidaGraph. The subsequent plot of k_{obs} verses pH, were plotted using sigmoidal curve fit and GraphPad Prism 4 (GraphPad Software). From this plot the pKa of sulphydryl group at C15 was determined.

The atomic coordinates and structure factors for OhrR C15S (1Z91) and the OhrR-*ohrA* complex (1Z9C) have been deposited in the Protein Data Bank, Research Collaboratory for structural Bioinformatices, Rutgers University (http://www.rcsb.org).

All figures were made using Swiss-Pdb Viewer (Guex and Peitsch, 1997), Grasp (Nicholls et al., 1991) and O and rendered in POV-Ray (www.povray.org).

Results and Discussion

Structure of reduced OhrR

The structure of OhrR C15S, in which residue C15 is replaced by a serine, was solved by multiple anomalous wavelength dispersion (MAD) (Hendrickson, 1991) and refined to 2.50 Å resolution (Table 3. 1). This mutant displays wild type DNA binding affinity (Fuangthong and Helmann, 2002) but does not suffer from oxidation during purification and crystallization. The structure of reduced wild type OhrR was also solved by molecular replacement using the OhrR C15S structure as the search model, and refined to 3.40 Å resolution. The structure of the OhrR C15S is identical to that of reduced wild type. Hence, we shall refer to OhrR C15S as reduced OhrR and describe its structure only.

The asymmetric unit (ASU) of the reduced OhrR crystal contains one subunit, which forms a biologically relevant dimer with a crystallographic dyadic mate. The fold of OhrR is similar to that of all other MarR family members and consists of six α helices

and three β strands with topology: α1, (residues 14-36), α2 (residues 40-51), β1 (residues 54-57), α3 (residues 58-64), α4 (residues 68-81), β2 (residues 83-88), β3 (residues 95-101), α5 (residues 102-122) and α6 (residues 130-143) (Figures 3. 9). Residues 10-13 and 107-110 form single turn 3_{10} helices and the nonhelical nature of residue 111 causes a discontinuity in α5. Each subunit is divided into two functional domains: the dimerization domain, which involves the N and C termini, α1, α2, α5 and α6 and buries 5,033 Å accessible surface area, and the winged HTH (wHTH) DNA binding domain, which consists of β1, α3, α4, β2 and β3. Helices α1 and α2 also form an interface between these well defined functional domains.

Structure of the OhrR-*ohrA* operator complex

The OhrR-*ohrA* operator complex crystallized in two different space groups, $P3_121$ and $P2_1$. Initially the structure of the complex was solved by MAD methods at 3.2 Å resolution in space group $P3_121$. Subsequently the structure of higher resolution $P2_1$ crystal form was determined by molecular replacement using the $P3_121$ structure as the search model and refined to 2.64 Å resolution (Figures 3. 9A, B and 4. 10A, Table 1). The asymmetric unit of the $P3_121$ crystal form contains one OhrR dimer bound to a 29 base pair oligodeoxynucleotide that encompasses the –10 element and transcription initiation site of the *ohrA* promoter. In addition to the OhrR-DNA complex, a third OhrR subunit was found that forms a crystallographic dimer with a neighbouring subunit and makes nonspecific contacts to the phosphate backbone of the DNA (Figures 3. 11A).

Three OhrR-*ohrA* operator complexes are found in the asymmetric unit of the $P2_1$ crystal form. Superimposition of all DNA bound OhrR subunits from the $P3_121$ and $P2_1$ crystal forms, using all Cα atoms results in root mean square deviations (rmsd) ranging from 0.7 Å to 0.9 Å. By contrast the superimposition of the subunits of the reduced OhrR and the DNA bound conformers reveals rmsd ranging from 1.5 Å to 1.7 Å (Figures 3. 11B). These changes are confined primarily to local rearrangements of the wHTH motifs. In order to bind the *ohrA* operator, the wHTH motif of one of the subunits of the reduced OhrR dimer rotates ~25° and the tips of the wings translocate ~16 Å (measured from the Cα atoms of E91', where prime indicates residues from the second subunit). This rotation results in the increased distance between the N-termini of the recognition helices of DNA-bound OhrR but a shortening of the C-termini of these helices (Figures 3. 11B). Although the wings of OhrR reorient significantly upon DNA-binding, their intersubunit tip-to-tip distances remain ~67 Å. Additional superimpositions of DNA-bound OhrR monomers onto those of MarR, MexR, SlyA-like protein and SarR result in rmsds of 1.6 Å, 2.0 Å, 1.9 Å and 1.9 Å, respectively, underscoring the similarity but inherent structural plasticity of the MarR family (Figures 3. 11C).

Interactions between OhrR and *ohrA* operator DNA

In general the DNA binding site of OhrR is B form but is bent ~10° and undertwisted 1.4°, properties that result in the shortening of the end-to-end distance by 3.5 Å (Ravishanker et al., 1989). Significantly different features of the OhrR-bound operator DNA conformation are found about the -10 region (TACAAT) and its two fold related sequence that include substantial widening (from 11.0 Å to 15.5 Å) and deepening (from

4.0 Å to 6.0 Å) of the major groove that result from insertion of the recognition helix (α4) of the HTH motif (Figures 3. 12). Moreover, the introduction of the tip of the wing into the minor groove and consequential contact from residue R94 to thymine 7 results in the local overtwisting of base pair thymine 7:adenine 24' (prime indicates bases from nontemplate DNA strand) from 34.2° in canonical B-DNA to 41°.

In total, 44 residues of the OhrR dimer make 60 DNA contacts over a span of 22 nucleotides, whereby each subunit engages in nearly identical, two fold related interactions with the DNA (Figures 3. 13A). The protein-DNA contact region includes the major groove of the -10 box (and its two fold related sequence) and indicates that OhrR, and likely MarR and MexR, repress transcription by blocking access of RNA polymerase to this promoter element. Previous mutational studies on the -10 elements of OhrR, MarR and MexR regulated promoters revealed a greater than 10 fold loss in DNA binding affinity when this region is altered (Evans et al., 2001; Fuangthong and Helmann, 2002; Lim et al., 2002; Martin and Rosner, 1995) .

The recognition helices of the HTH motifs of the OhrR dimer bind two consecutive major grooves of the *ohrA* operator whereby the N-termini of each helix points towards the floor of the major groove to make base specific contacts (Figures 3. 11A, 4. 13A and B). The number of protein-DNA base contacts is fairly small and includes hydrogen bonds between the carbonyl oxygen of G69 and N4 of cytosine 18' and between the Oγ of S68 and N7 of adenine 10 and van der Waals contacts between the Cα of G69' and O4, C4, C5 and C7 of thymine 11 and between the S68 Oγ and C8 of adenine 10. The Oγ of T72

hydrogen bonds to the adenine 9 phosphate group and the side chains of residues K76 and R77 interact with the phosphates of thymine 11' and adenine 15', respectively. Conserved MarR family residue D67 makes water-mediated base contacts to N6 and N7 of adenine 19'. Additional major groove contacts include those between the Oγ of T70 and N7 of adenine 17', Oγ of S68' and the phosphate moiety of adenine 8' and the side chains of Y65 and L66' and the phosphate backbones of adenine17' and cytosine 18, respectively.

The second DNA binding element, the wing, is composed of strands β2 and β3 and their connecting loop (residues 89-94) and is quite long, extending seven base pairs (~24 Å) upstream and downstream from the α4 recognition helices (Figures 3. 11A). A wing-DNA minor groove hydrogen bond is formed between the NH of R94 and the O2 and O4' of thymine 7 (Figures 3. 13C). The former contact effectively discriminates against purines at this position. The minor groove position of R94 is buttressed by its electrostatic interaction with the carboxylate side chain of D92 and a phosphate contact from the guanidinium side chain of R86 of β2. Residues R94 and D92 are strongly conserved amongst MarR family proteins and a similar pyrimidine-R94-D92 interaction network, locked into place by an R86-DNA backbone contact, will likely be found throughout the family (Figures 3. 10B). The importance of R94 in MarR family function is underscored by mutational studies on MarR and MexR whereby substitution of this arginine with alanine or cysteine resulted in a 10 fold loss in DNA binding affinity (Alekshun et al., 2001; Saito et al., 2003). Also contributing to the DNA binding affinity of OhrR is the positive dipole moment originating from the N-terminus of helix α3,

which is aimed at the phosphate backbone of base pairs 7 and 8 (Figures 3. 11A). Additional wing-minor groove hydrogen bonds include a water-mediated interaction between the guanidinium side chain of R88 and O2 of thymine 8 and O4' of adenine 9 and a direct hydrogen bond between the main chain amide of E93 to DNA phosphate backbone of thymine 26' (Figures 3. 13A and D). Residue D92' also engages in van der Waals contacts to C5' of adenine 25 and the phosphate backbone of thymine 26.

In addition to the wHTH motif, OhrR contains a third DNA binding element, a Helix-Helix motif (HH), which is composed helices α1 and α2 and their connecting residues (Figures 3. 11A). Helix α1 of the HH motif crosses the pseudodyad axis to interact with the DNA. Residues Y19, K30 and T39 of this element engage in hydrogen bonds with the phosphate groups of adenines 17', adenine 19 and thymine 16', respectively, whilst residues R23 and K27 approach the phosphate backbone of cytosine 18. Moreover, the positive macrodipole moment of α2 is directed toward the DNA phosphate backbone surrounding the pseudodyad of the operator. Thus, each subunit of the OhrR dimer locks onto the *ohrA* operator by an extended wing, a HTH motif and the HH interaction motif, a newly described element that is likely common to all MarR family members.

The reactive site and organic hydroperoxide specificity

The *B. subtilis* OhrR cysteine residue, C15, which is conserved amongst all described OhrR proteins, is oxidized to sulphenic acid by OHPs such as *tert*-butyl and cumene hydroperoxides (Fuangthong et al., 2001; Fuangthong and Helmann, 2002). The structure of OhrR reveals that C15 is located at the N-terminus of helix α1 suggesting

that besides its role in DNA binding, the HH motif play a key role in the peroxidation-mediated induction mechanism of OhrR. In both the reduced and DNA-bound structures, the side chain of residue 15 is hydrogen bonded identically to the side chain hydroxyls of Y29' and Y40', which are located on helices α1' and α2', respectively, and a solvent molecule, thereby attesting to the solvent accessibility of the sensor (Figures 3. 14A and B). In OhrR, unlike the oxidative stress response transcription regulators, OxyR (Choi et al., 2001) and Spx (KJ Newberry, personal communication) no basic residues are found within 7 Å of the OhrR C15 side chain and thus, other chemical or structural features must influence its high reactivity. A novel and most likely candidate to explain a lower pK_a of C15 is the positive macrodipole (Hol et al., 1978; Wada, 1976) of helix α1, at which C15 is located. Such positioning should effectively lower the pK_a of the side chain of the sensor to allow the more facile formation of the reactive, nucleophilic thiolate at physiologically relevant pH. Indeed, determination of the pK_a of a cysteine side chain that is located at the N-terminus of an α helix of a model peptide and engaged in hydrogen bonding revealed a decrease from 8.83 to 7.20 (Kortemme and Creighton, 1995). Moreover, the presence of two hydrogen bond donors from the side chains of residues Y29' and Y40' would stabilize the negatively charged thiolate in of reduced form of OhrR protein (Figure 3. 14A).

In accord with our hypothesis that C15 is appropriately reactive at physiological pH, the pK_a of the C15 sulphydryl group was determined using site directed fluorescence labelling and the bimane derivative, (2-pyridyl)-dithiobimane (PDT-Bimane) (Carlsson et al., 1978; Mansoor and Farrens, 2004). Reaction between the cysteine thiolate and PDT-

Bimane results in the release of pyridyl-2-thione, which can be monitored spectroscopically at 343 nm and the $T_{1/2}$ of half maximal release or reactivity can be calculated (Figure 3. 14C). From this the $T_{1/2}$ of half maximal release or reactivity over a pH range from 4.5 to 8.5 plotted against pH yields the pK_a of the sulphydryl group. The pKa of OhrR resides C15 is ~5.2 (Figure 3. 14D). Thus, the N-terminally located C15 is in the thiolate form at physiological pH and should be ready to react with OHPs.

In addition to their role in stabilizing the reactive site cysteine side chain, residues Y29 and Y40 enforce a steric constraint upon the OhrR reactive site such that oxidation of the C15 side chain to sulphenic acid would result in steric clash with their phenolic side chains if no compensatory conformational change occurred (Figure 3. 15). In support of this supposition is the finding that substitution of Y29 or Y40 or both to the much smaller alanine results in OhrR proteins that can bind the *ohrA* operator but fail to derepress the *ohrA* gene in response to oxidative stress (S. Soonsarga, personal communication), i.e., a sulphenic acid side chain can be accommodated without steric clash. The required conformational change of Y29 or Y40 can be readily envisioned as the first domino to fall as nearby DNA contacting residues, e.g., Y19 and its two fold mate, must too reposition their side chains to avoid subsequent steric clash (Figure 3. 15A). Ultimately, the ensuing side chain conformational cascade would dislodge the DNA binding domains from the *ohrA* operator.

ROS are toxic and must be removed rapidly and hence a variety of mechanisms are in place to detoxify them. The OhrR-*ohrA* system is induced by and detoxifies OHPs

specifically. Although OhrR is able to respond to hydrogen peroxide, in *B. subtilis* other enzymes such as a major vegetative catalase (KatA) and alkyl hydroperoxide reductase, which respond to much lower levels of H_2O_2 than OhrR does, constitute the most important pathways for detoxifying H_2O_2 *in vivo* (Fuangthong and Helmann, 2002). The structure of OhrR provides a chemical rationale for the preference of organic hydroperoxides over hydrogen peroxide as the *in vivo* inducers. The presence of a long continuous strip of aromatic and nonpolar residues on helix α2 (Y_{43}LALLLLW_{50}, Figures 3. 10 B and 14B), which is not conserved amongst the MarR family but is strongly conserved amongst all OhrR proteins, provides a solvent exposed, ~12 Å long hydrophobic landing pad for such lipophilic oxidants. Recent biochemical studies by the Mongkolsuk group (Klomsiri et al., 2005) strongly support our structural inferences. The natural product, linoleic acid hydroperoxide (LOOH), which contains an 18 carbon unsaturated alkyl tail, that would span the entirety of this hydrophobic patch, induces expression of the *X. campestris ohr* by 80 fold through its oxidation of OhrR. By contrast the smaller OHP, *tert*-butyl hydroperoxide, induces *ohr* expression by only 10 fold and hydrogen peroxide does not induce at physiologically relevant concentrations. Moreover, the concentration of LOOH required to completely inhibit the binding of OhrR to operator DNA was 100-fold lower than that for *tert*-BOOH (Klomsiri et al., 2005).

Chimeric nature of the wHTH motif of the MarR family

A striking feature of the wHTH motif of OhrR is its long wing. An overlay of the OhrR wHTH onto the wHTH motifs of the bacterial transcription regulators BmrR (Zheleznova et al., 1999) and CRP (Schultz et al., 1991) highlights this difference as well as the

structural conservation of the prokaryotic HTH motif (Figures 3. 14C). By contrast, the superimposition of the OhrR wHTH motif onto that of the eukaryotic transcription regulator RFX1 (Gajiwala et al., 2000) and histone H5 (Ramakrishnan et al., 1993) reveals differences in the lengths and orientations of the HTH motifs but remarkable similarity in the length and orientation of the wings. Although a similar extended wing has been observed in a few other bacterial proteins, such as MecI, BlaI and ArsR/SmtB families (Cook et al., 1998; Garcia-Castellanos et al., 2004; Klomsiri et al., 2005), the DNA binding mode of their wings resembles that of their shorter winged bacterial counterparts, whilst that of OhrR resembles more closely those of the eukaryotic wHTH motifs. Hence, the wHTH motif of OhrR is a structural chimera that combines the canonical HTH motif of prokaryotes and the wing of eukaryotes with a novel element, the Helix-Helix motif, to effect high affinity DNA binding. A similar DNA binding mechanism is anticipated for the entire MarR family.

Acknowledgements

We thank Drs. K. J. Newberry and M. A. Schumacher for their outstanding crystallographic help and scientific discussions and Dr. C. Ralston and other beamline scientists at ALS BL 8.2.1 for help with data collection. We thank S. E. Mansoor for help with PDT-Bimane experiments. We also thank other members of the Brennan laboratory for helpful comments. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory. This work was supported in part by the

National Science Foundation (J.D.H.), the Richard T. Jones Professorship of Structural Biology (R.G.B.) and the Tartar trust fellowship (M.H.).

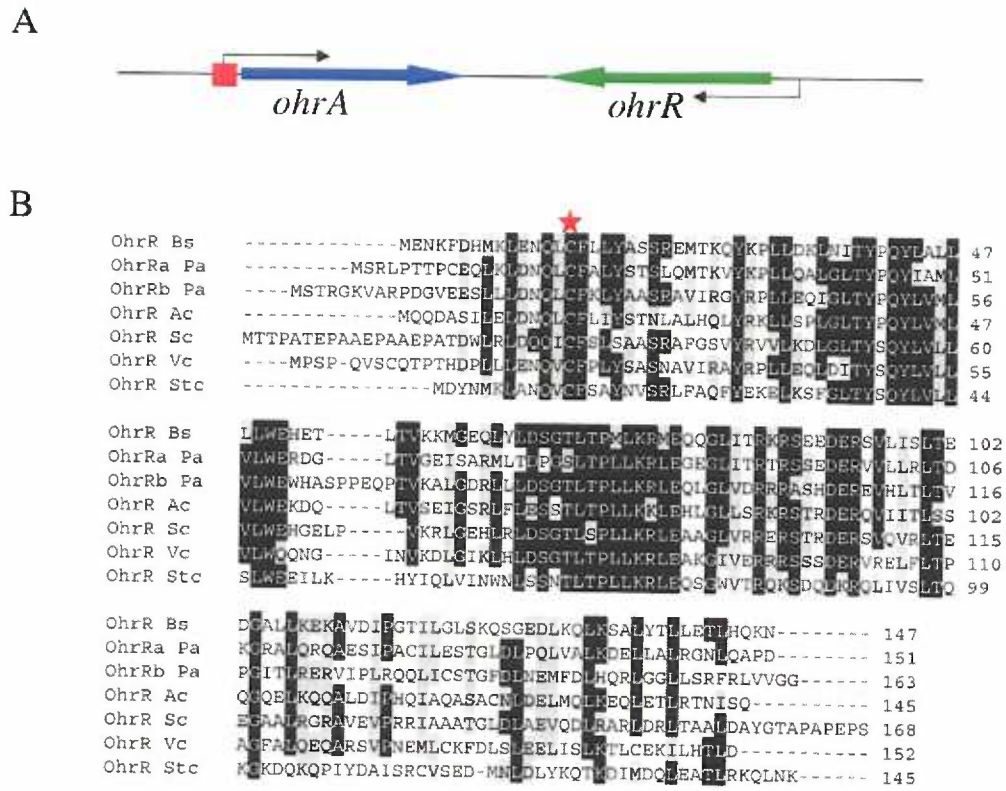


Figure 3. 1 *ohrR* encodes a MarR-family repressor of *ohrA*. **A**, Schematic showing the operator region of *ohrA* and *ohrR*. Red box indicates the operator site for the OhrR protein in the *ohrA* promoter. **B**, Alignment of OhrR homologues. The abbreviations used are as follow (strain; GenBank accession number): OhrR Bs (*B. subtilis*;E69857), OhrRa Pa (*P. aeruginosa* PAO1; D83290), OhrRb Pa (*P. aeruginosa* PAO1; G83292), OhrR Ac (*Acinetobacter* sp. strain ADP1; CAA70318), OhrR Sc (*Streptomyces coelicolor*; CAB87337); OhrR Vc (*Vibrio cholerae* group O1 strain N16961; B82389), and OhrR Stc (*Staphylococcus sciuri* strain ATCC 29062). The amino acid sequences were aligned (using CLUSTALW) and conserved residues highlighted using the BoxShade utility. A conserved cystein residue is denoted by a red star. Figure adapted from (Fuangthong et al., 2001)

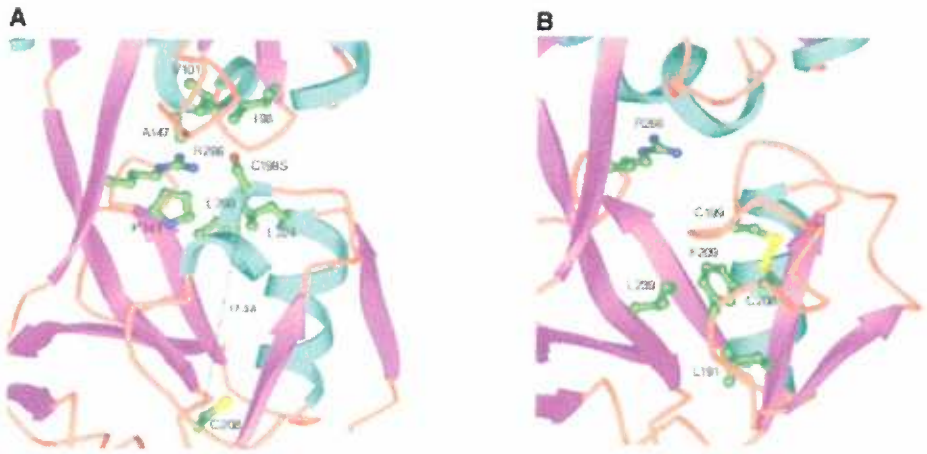


Figure 3. 2 Structure of the OxyR Regulatory Domain. The ribbon diagrams of the OxyR monomers in the reduced (**A**) and oxidized (**B**) forms are shown with the redox-active cysteines Cys-199 (C199S in the reduced form), Cys-208 and the neighboring residues (including Arg-266) in a ball-and-stick representation. Figure is reproduced from (Choi et al., 2001)

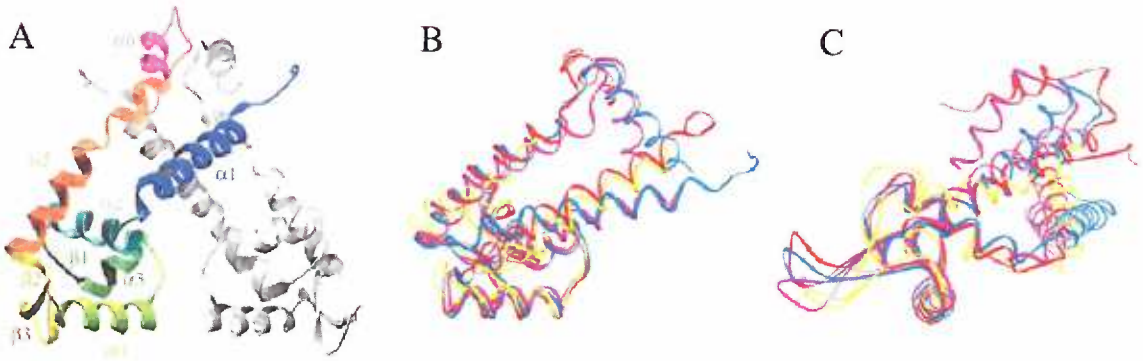


Figure 3. 3 Structures of MarR family proteins. **A**, Overall structure of the MarR dimer, (Alekshun *et al.*, 2001). **B**, Overlays of the structures of MarR (PDB ID code 1JGS), MexR (1LNW), SlyA-like (1LJ9), and SarR (1HSJ) monomers, colored in red, blue, yellow and purple, respectively. The monomers were superimposed using the main chain atoms of the recognition helix, α4, as depicted in A. **C**, The superimposed structure rotated by 90° relative to B.

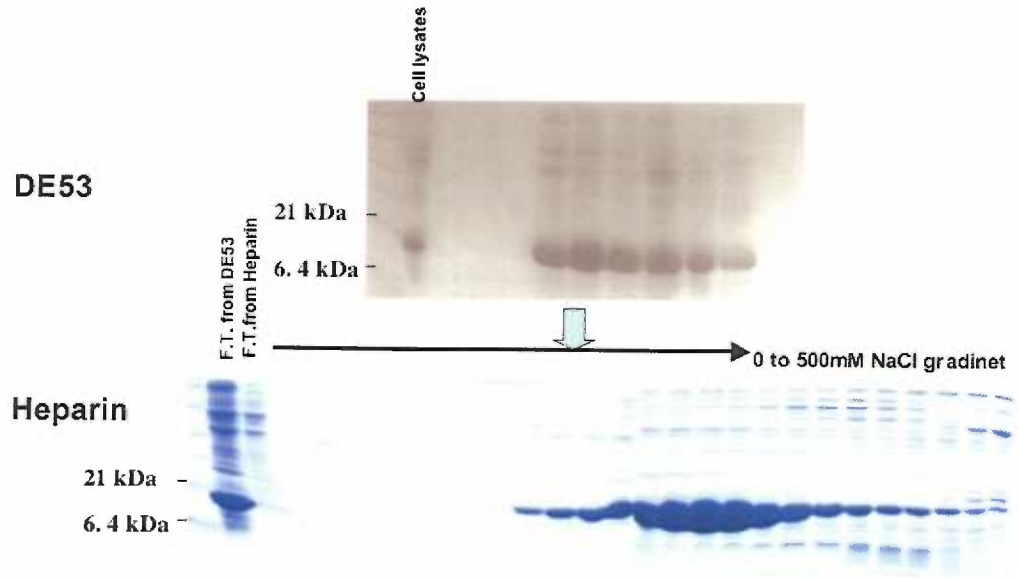


Figure 3. 4 Purification of *B. subtilis* recombinant OhrR protein. Cell lysates of *B. subtilis* OhrR protein overexpressed in *E. coli* were applied onto a DE 53 anion exchange column. The flow-thorough fractions containing OhrR protein were pooled together then applied onto a heparin sepharose column. Pure OhrR protein was eluted using NaCl gradient from 0 to 500 mM.

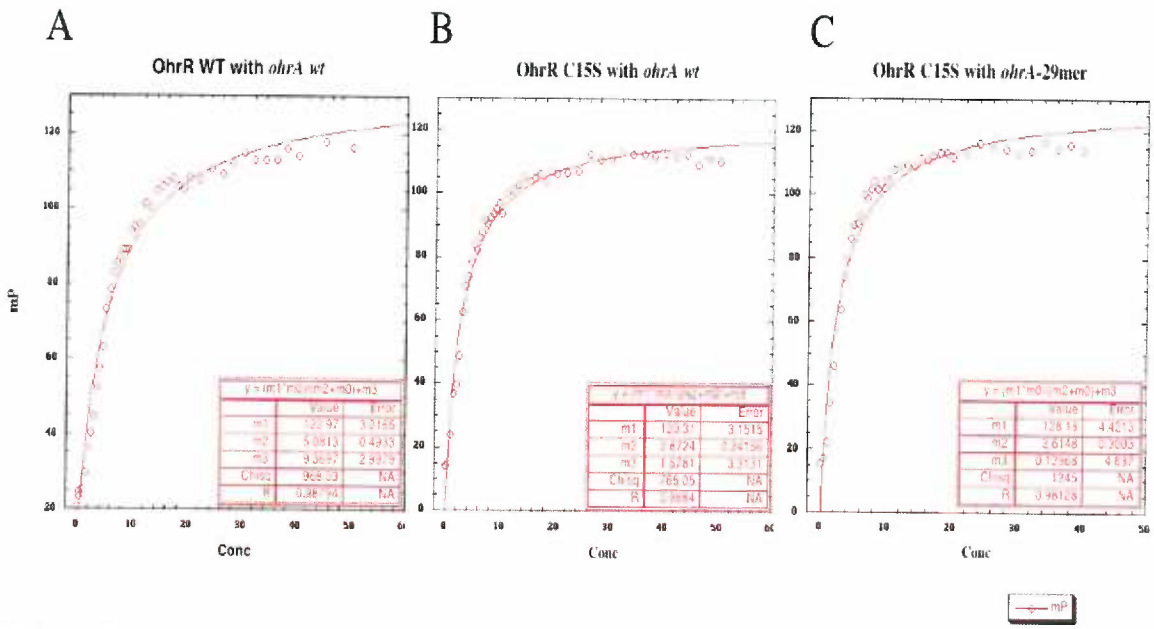


Figure 3. 5 OhrR equilibrium binding to operator DNA. Millipolarization (mP) is plotted against the concentration of OhrR protein. **A**, OhrR wild type binding to *ohrA* wild type, **B**, OhrR C15S to *ohrA* wild type and **C**, OhrR C15S to *ohrA*-29mer.

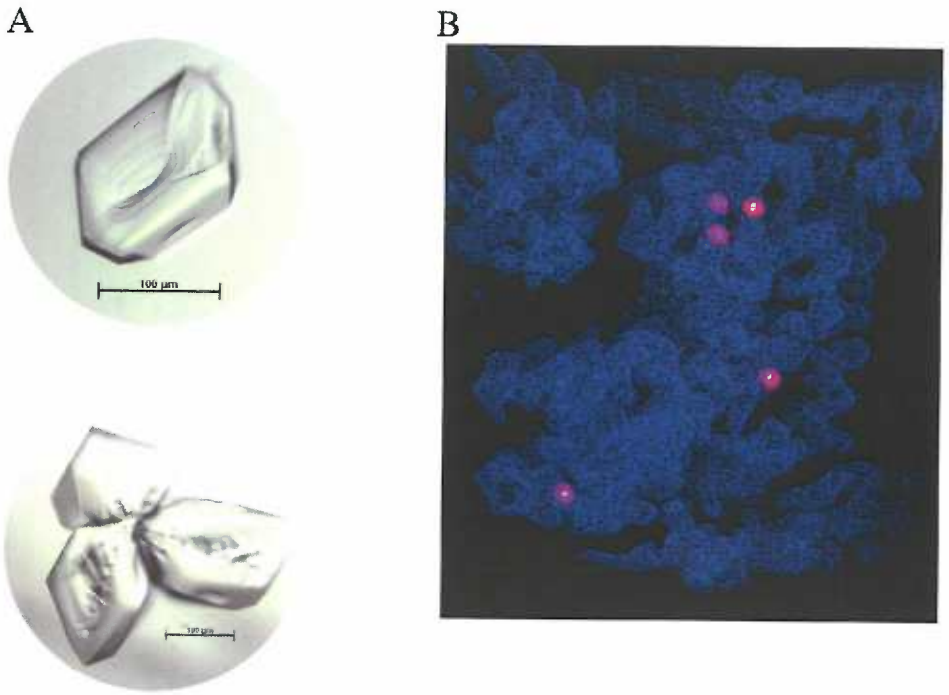
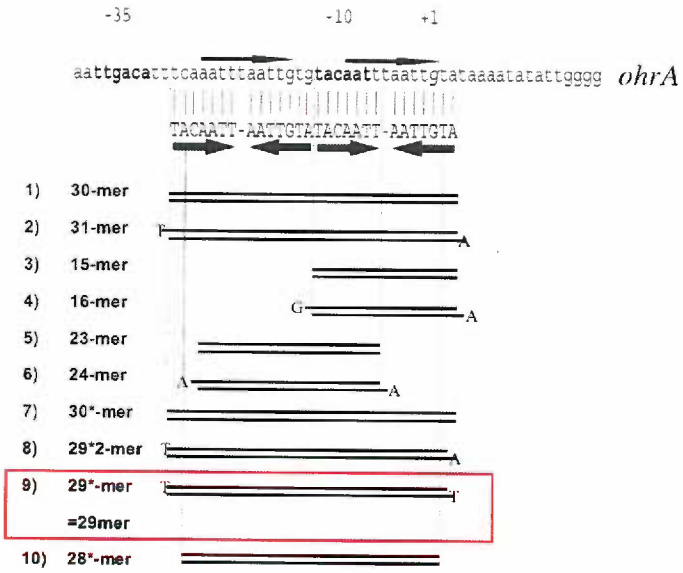


Figure 3. 6 **A**, Pictures of reduced OhrR crystals. **B**, Initial electron density map of the reduced OhrR C15S, after density modification. The map is calculated to 3.0 Å resolution and contoured at 1σ. Selenomethionine sites are shown as red spheres.

A



B

ohrA-wt F -C ATT TCA AAT TTA ATT GTG TAC AAT TTA ATT GTA TAA A
G TAA AGT TTA AAT TAACAC ATG TTA AAT TAA CAT ATT T

ohrA-29mer F -C ATT TAC AAT TTA ATT GTA TAC AAT TTA ATT GTA TAA A
G TAA ATG TTA AAT TAA CAT ATG TTA AAT TAA CAT ATT T

Figure 3. 7 **A**, *B. subtilis ohrA* promoter region (Fuangthong and Helmann, 2002) and the deoxyoliogonucleotides that were used in OhrR-*ohrA* complex crystallization trials. The red box indicates the pieces used in OhrR-*ohrA* complex structure determination. –35 and –10 boxes and +1 transcription initiation sites are indicated above the sequences. **B**, Shown are the *ohrA* wild type promoter sequences and the sequences of *ohrA*-29 mer used in OhrR-*ohrA* structure determination. Mutations in ohrA-29mer used enhance the palindromic nature of the binding site are colored in red.

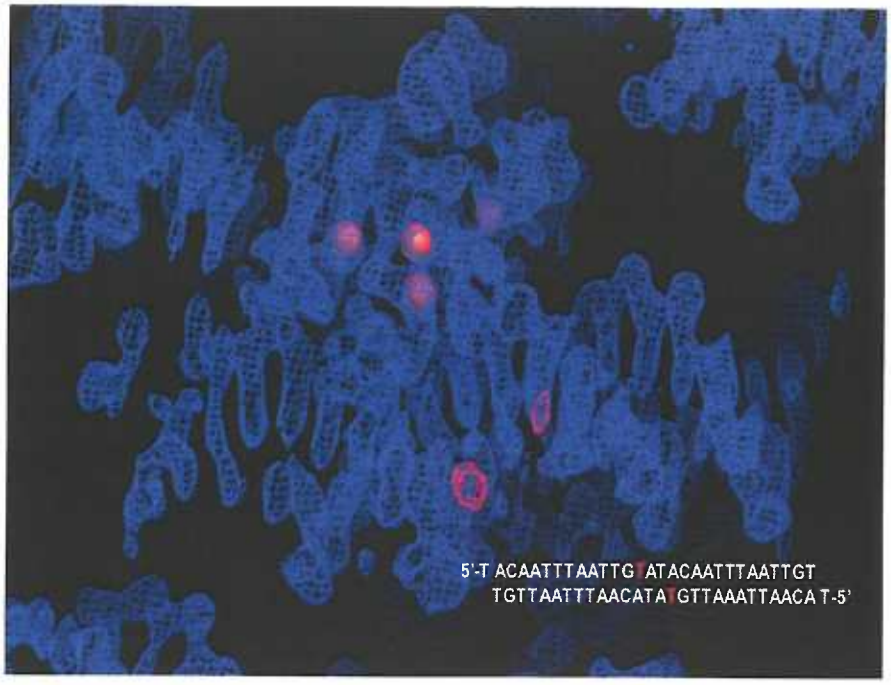
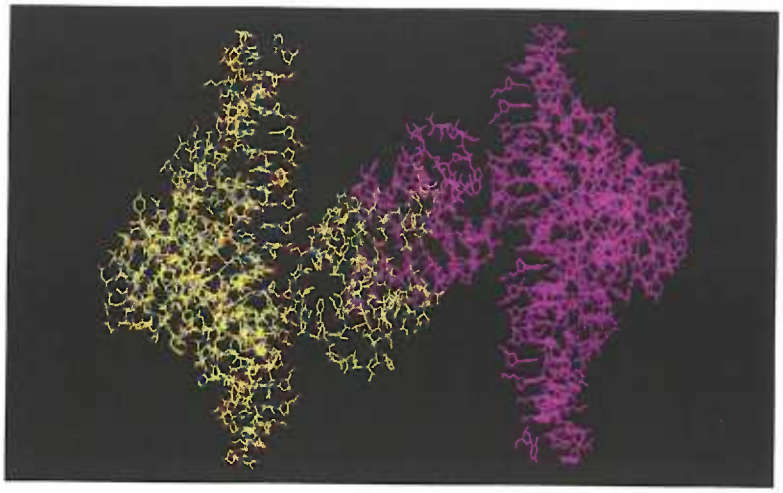


Figure 3. 8 Electron density map of the OhrR-DNA complex from $P3_121$ crystal form. The map was calculated after density modification to 3.2 Å resolution and contoured at 1σ. The difference fourier map showing 5′-iododeoxyuracil sites are shown in red mesh is calculated to 4 Å resolution and contoured at 4 σ. Selenomethionine sites are shown as red spheres.

A



B

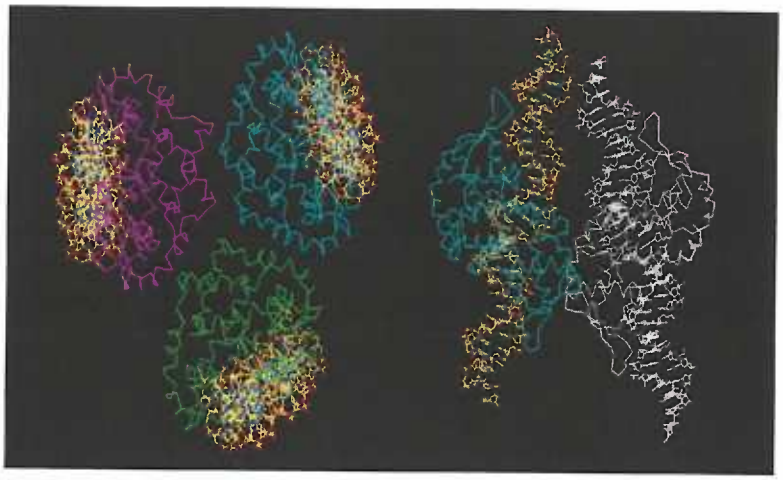


Figure 3. 9 OhrR-*ohrA* complexes in the asymmetric unit of the $P3_121$ (**A**) and $P2_1$ (**B**) crystals forms, respectively. **A**, Crystallographic symmetric mates are shown in purple. **B**, Left, each OhrR dimer is shown as purple, green and blue. Right, the crystallographic symmetric mate is shown in white.

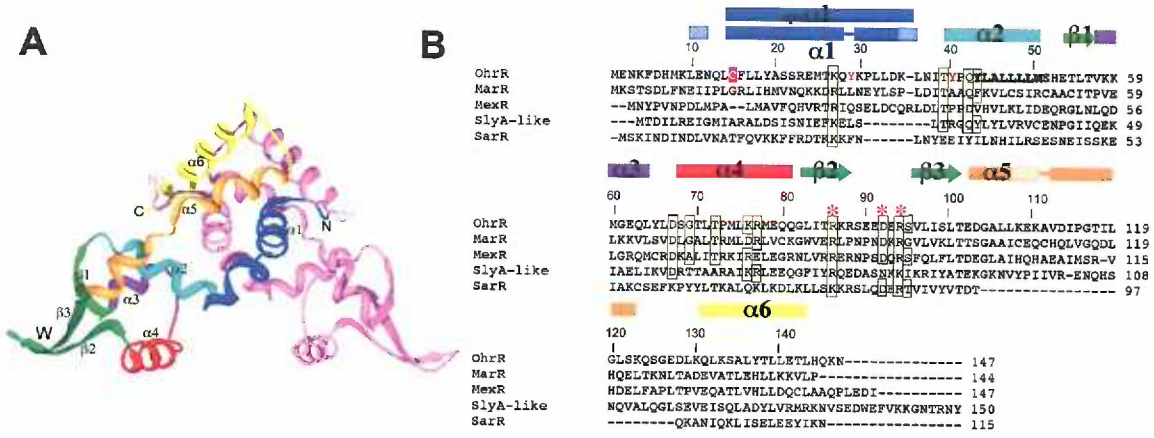


Figure 3. 10 Structure of reduced OhrR. **A**, OhrR is shown as a ribbon diagram with the N and C termini of both subunits and secondary structural elements of one subunit coloured and labelled. The dyadic mate is coloured pink. One wing is labeled with a "W". **B**, Sequence alignment of structurally determined MarR family proteins. The secondary structure elements of OhrR and DNA-bound OhrR are indicated above the sequence by arrows (β strands), solid rectangles (α helices) and hatched rectangles (3_{10} helices). The colour of each secondary structure matches those of Figures 1a and 2a. Reactive site residue C15 is encased in a red box and Y29 and Y40 are coloured red. DNA contacting residues of the wing are denoted by red asterisks. Selected homologous/identical residues that interact with DNA are boxed in gold. An eight-residue stretch of nonpolar/aromatic residues that is likely involved in organic hydroperoxide binding is bolded.

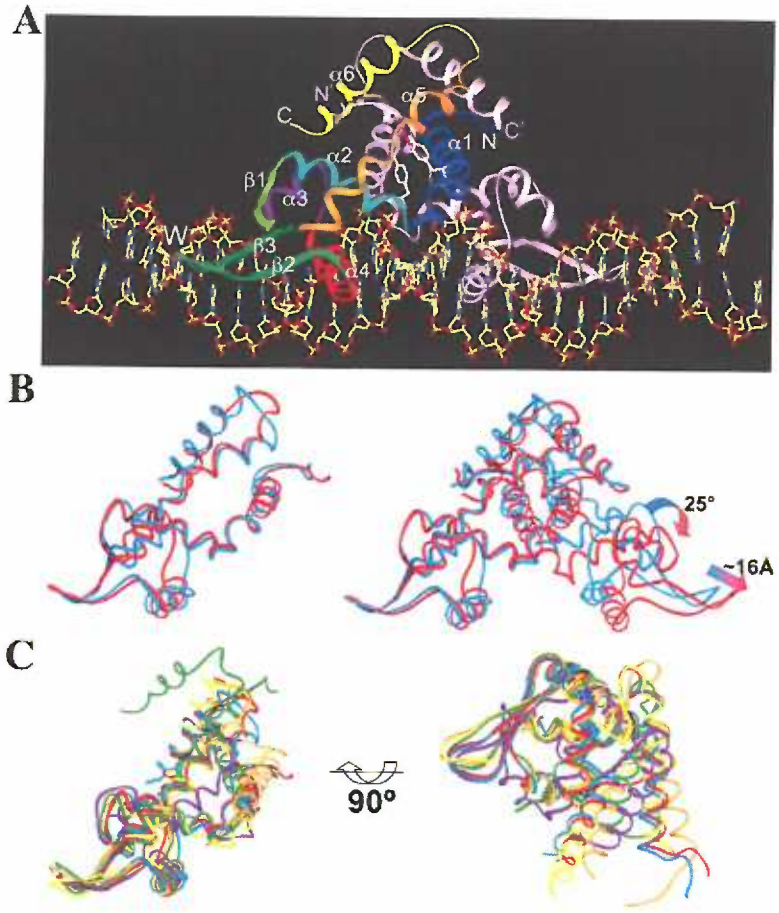
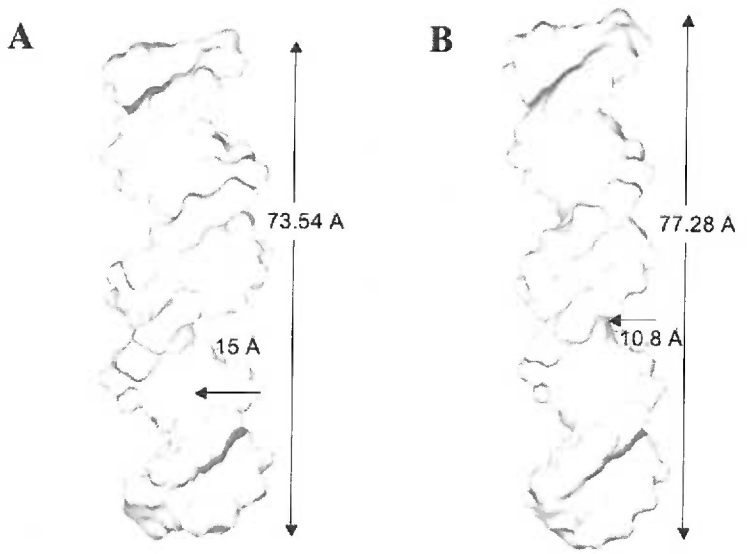


Figure 3. 11. Structure of the OhrR-*ohrA* operator complex. **A**, Ribbon diagram of OhrR dimer coloured and labelled as in Figure 1. The *ohrA* operator is shown as a stick model with carbons coloured yellow; nitrogens, blue; oxygens, red and phosphorus, orange. Reactive site residue C15 and interacting tyrosine residues Y29 and Y40 are coloured in pink and white, respectively. **B**, Superimpositions of the subunits (left) or dimer (right) of the DNA-bound (red) and reduced OhrR (blue), using all corresponding Cα atoms in one subunit. The curved arrow indicates the rotation of the second subunit of the DNA-bound OhrR dimer relative to that of the reduced OhrR dimer. The straight arrow reveals the translation of the wing as measured between the Cα atoms of residue

E91′. C, (Left) Ribbon diagrams of the optimized superimpositions of OhrR-DNA (red), reduced OhrR (blue), MarR (green, PDB ID 1JGS), MexR (orange, 1LNW), SlyA-like protein (yellow, 1LJ9), and SarR (purple, 1HSJ) using all corresponding Cα atoms. (Right) The overlays rotated by 90° along the x axis.



C

	OhrR-DNA complex	B-DNA
Twist	34.6°	36°
Bend	9.8°	0°
No. of n. t. / turn	10.4	10
Rise / n. t.	3.23 Å	3.36 Å

Figure 3. 12 Major groove deepening of *ohrA* by OhrR binding. Comparison of the OhrR-bound *ohrA* 22 bp operator (**B**) to a B-DNA strand (**A**). **C**. Selected parameters for B-DNA and *ohrA* structure analysis calculated using Curves.

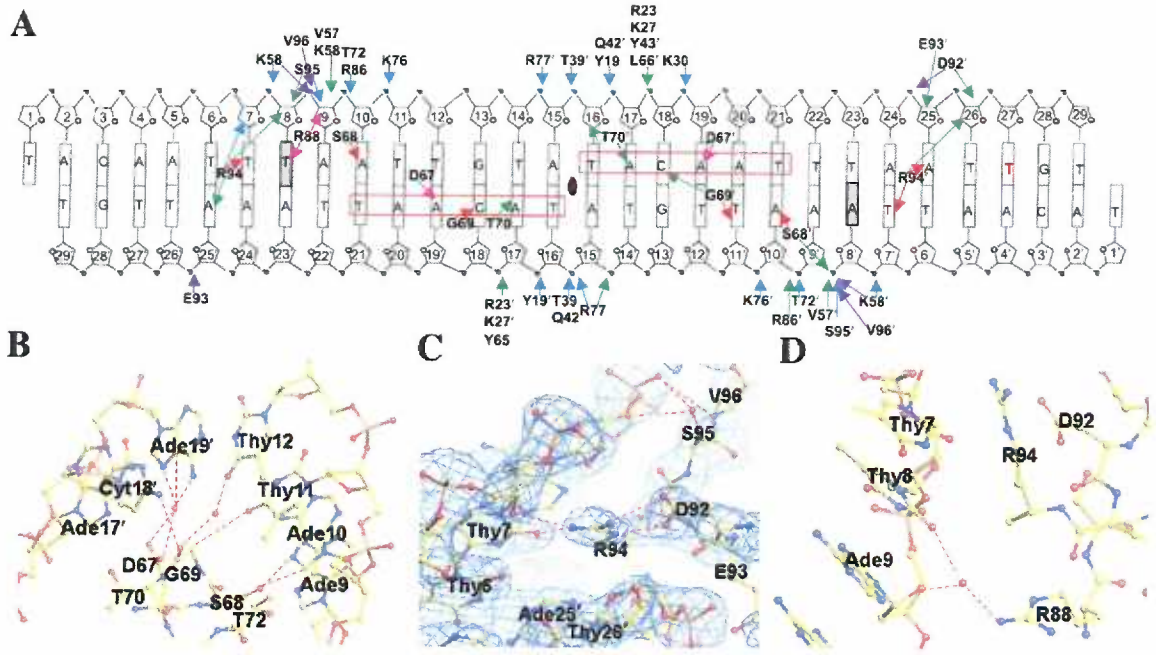


Figure 3. 13 OhrR-*ohrA* operator binding. **A**, Schematic showing all OhrR-DNA contacts. The bases are labelled and shown as rectangles. The –10 element, bps 16-21, is boxed in red as is its two fold related sequence (bps 10-15). The transcription start site (bp 27) is bolded in red. Thymindine 8 and adenine 8', which are not dyad-related, are shaded. Red arrows indicate base specifying protein-DNA contacts; pink arrows, water mediated protein-DNA hydrogen bonds; blue arrows, side chain-DNA phosphate backbone interactions; purple arrows, main chain-DNA phosphate backbone interactions; green arrows, van der Waals contacts. Primed residues belong to the second subunit. Primed deoxyriboses denote the other DNA strand. The pseudodyad axis relating the left and right halves of the DNA is indicated by an oval between bps 15 and 16. **B**, Major groove contacts. **C**, Minor groove contacts. The $2F_o$-F_c simulated-annealing composite omit electron density, contoured at 1σ, is shown as blue mesh. **D**, Water mediated minor

groove-β2 contacts. In B, C and D interacting residues and nucleotides are labelled, shown as balls and sticks and coloured as in Figure 2A. Solvent molecules are shown as red spheres. Hydrogen bonds are depicted by dashed lines.

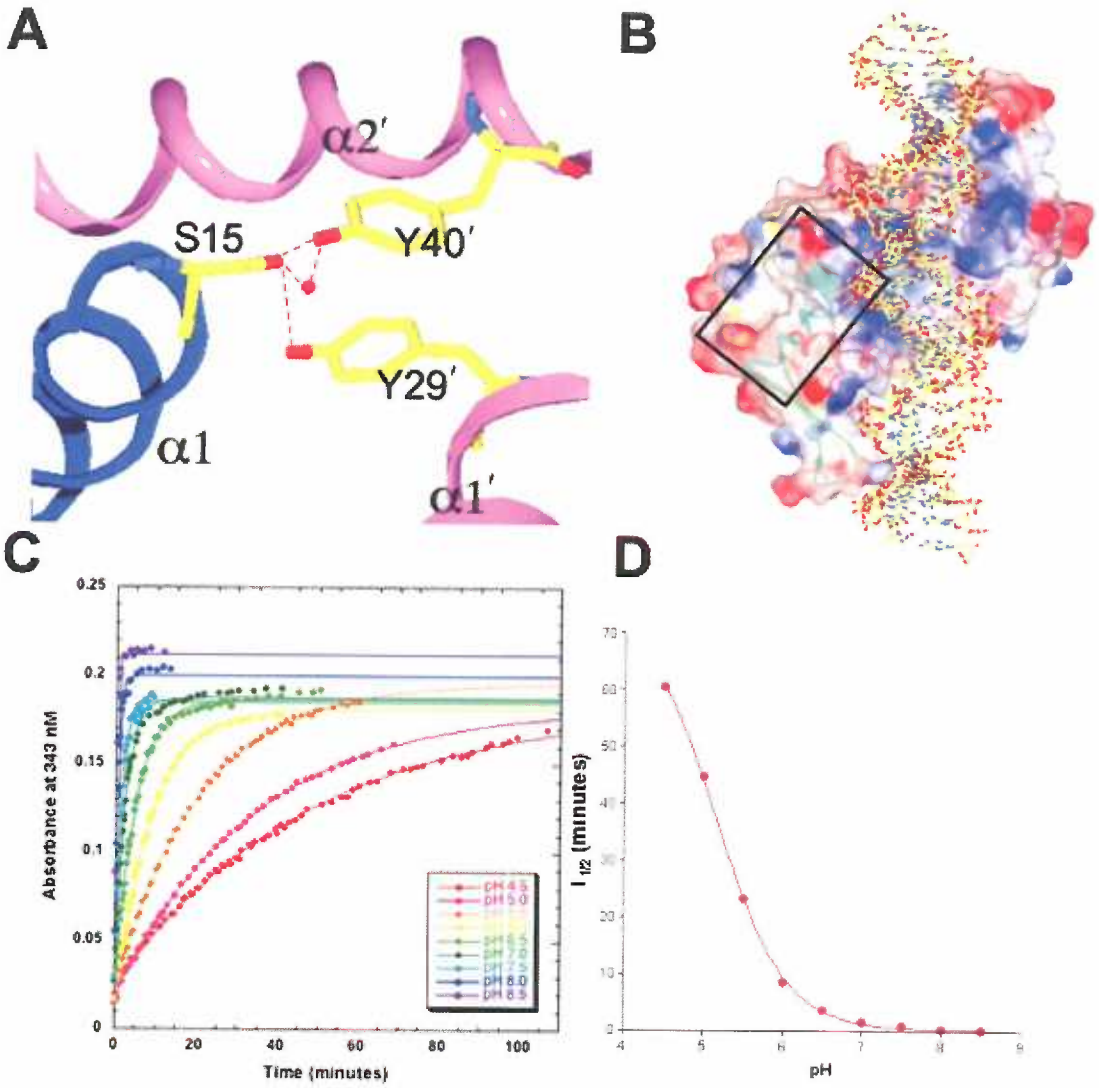


Figure 3. 14 The OhrR reactive site. **A**, Close up of the OhrR reactive site. Residues, C15S, Y29′ and Y40′, are shown as sticks where oxygen, nitrogen and carbon atoms are coloured red, blue and yellow, respectively. The red sphere is a solvent molecule that is hydrogen bonded (dashes) to the side chain hydroxyls of residues 15 and 40′. **B**, The electrostatic surface potential of DNA-bound OhrR. Blue, red and white colours indicate positive, negative and neutral electrostatic regions, respectively. The site of oxidation and the hydrophobic patch of helix α2

leading to C15 is boxed. Residue 15 is depicted by green van der Waals spheres. The protein structure is shown as a ribbon diagram and coloured as in Figure 2A. The DNA is depicted by sticks with atom types coloured as in Figure 2A. **C**, The pH dependence of OhrR labelling by PDT-bimane. Plotted is the absorbance at 343 nm at pHs ranging from 4.5 to 8.5 resulting from the release of by pyridyl-2-thione after the reaction of OhrR C15 and PDT-bimane (y axis) against time in minutes along the x axis. **D**, The pH dependence of the $T_{1/2}$ (time of half maximal reactivity) of the OhrR C15-(PDT-bimane) reaction and release of pyridyl-2-thione. The $T_{1/2}$ is plotted on the y aixs against the pH (x axis). From this plot the pK_a of C15 sulphydryl group is ~6.4.

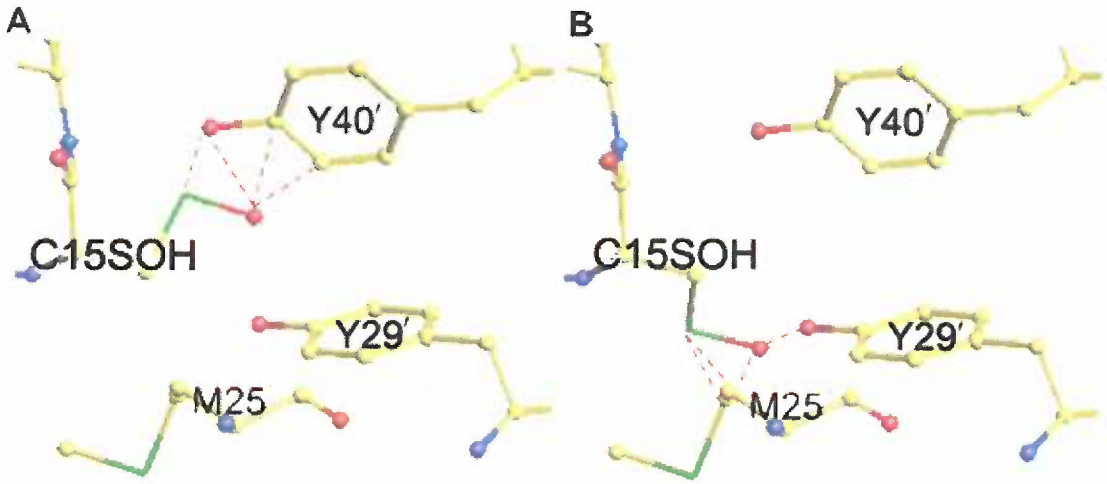


Figure 3. 15. The OhrR reactive site in which a cysteine sulphenic acid side chain has been modelled at residue C15 (C15SOH). Dashed lines indicating steric clash between C15SOH and neighbouring residues in the reactive site of the reduced form, show interatomic distances less than 2.2 Å. **A**, C15SOH-Y40′ steric clash and **B**, C15SOH-Y29′ and M25′ steric clash. Other possible conformations of C15SOH result in steric clash with the backbone amide group of residue F16. These observations suggest strongly that, upon oxidation by organic hydroperoxide, the formation of C15SOH in the reactive site must initiate structural changes in OhrR that alleviate steric clash and lead to dissociation from the *ohrA* operator site.

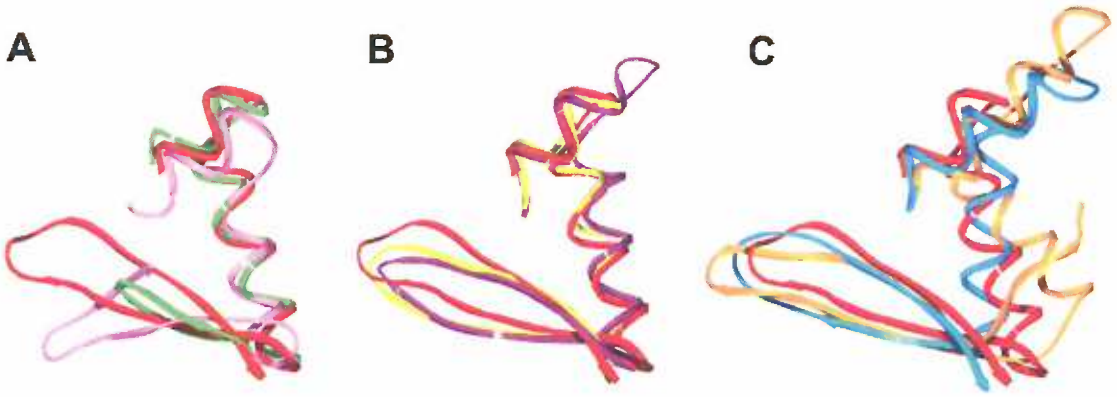


Figure 3. 16 Chimeric nature of the OhrR wHTH motif. Optimal superimpositions of the Cα atoms of the wHTH motifs of **A,** OhrR (red), CRP (green, 1CGP), BmrR (pink, 1EXJ), **B,** OhrR (red), MecI (purple, 1sax), SmtB (yellow, 1smt) and **C,** OhrR (red), human regulatory factor X (orange, 1DP7) and the globular domain of histone H5 (blue, 1HST).

Table 3. 1 Selected crystallographic data and statistics

Data analysis reduced OhrR	Native	Seleno-reduced OhrR MAD		
Wavelength (λ)	0.9794	0.9794	0.9184	0.9796
Resolution (Å)	55.26-2.50	84.81-3.00	84.81-3.00	84.81-3.00
Overall R_{sym} [a]	0.09 (0.10) [b]	0.06 (0.23)	0.07 (0.22)	0.05 (0.25)
Overall I/σ(I)	7.0 (4.8)	7.1 (3.3)	7.3 (2.9)	7.8 (1.9)
Total Reflections (#)	27682	33926	37323	33859
Unique Reflections (#)	4514	2998	3052	2986
Completeness (%)	99.8 (99.8)	98.2 (89.9)	99.1 (88.9)	98.2 (89.6)
Overall Figure of Merit[c]			0.58	

OhrR-*ohrA* operator	Native	Seleno-OhrR-*ohrA* MAD		
Space Group	$P2_1$		$P3_121$	
Wavelength (λ)	0.9794	0.9794	0.9184	0.9796
Resolution (Å)	55.26-2.50		91.29-3.20	
Overall R_{sym} [a]	0.09 (0.10) [b]	0.05 (0.09)	0.07 (0.08)	0.06 (0.08)
Overall I/σ(I)	7.0 (4.8)	9.9 (8.5)	9.5 (8.6)	9.1 (8.8)
Total Reflections (#)	27682	50872	66130	62251
Unique Reflections (#)	4514	10711	11684	11251
Completeness (%)	99.8 (99.8)	95.8 (77.8)	99.8 (98.7)	98.2 (88.9)
Overall Figure of Merit[c]			0.53	

Refinement Statistics	reduced OhrR		OhrR-*ohrA* operator	
	Native	MAD	Native	MAD
Resolution (Å)	77.10-2.50	84.80-3.00	59.65-2.64	91.00-3.20
Space group	$P4_121$	$P4_121$	$P2_1$	$P3_121$
R_{work}/R_{free}(%)[d]	22.8/29.5	27.5/28.3	21.5/27.9	29.9/33.6
Bond angles (°)	1.20	1.32	1.20	1.40
Bond lengths (Å)	0.007	0.01	0.007	0.01
B values (Å^2)				
Average B values (Å^2)	37	34	45	39
OhrR protein			49	32
ohrA oligodeoxynucleotide			48	45
Ramachandran analysis				
Most favoured (%/#)	87.1/108	84.6/104	89.1/665	71.6/239
Add. allowed (%/#)	10.5/13	11.4/14	10.2/76	24.0/80
Gen. allowed (%/#)	1.6/2	1.6/2	0.6/5	3.9/13
Disallowed (%/#)	0.8/1	2.4/3	0	0.6/2

[a] $R_{sym} = \Sigma\Sigma|I_{hkl}-I_{hkl(j)}|/\Sigma I_{hkl}$, where $I_{hkl(j)}$ is the observed intensity and I_{hkl} is the final average intensity value.

[b] values in parentheses are for the highest resolution shell.

[c] Figure of Merit = $<|\Sigma P(\alpha)e^{i\alpha}/\Sigma P(\alpha)|>$, where α is the phase and P(α) is the phase probability distribution.

[d] $R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$ and $R_{free} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$; where all reflections belong to a test set of 10% randomly selected data for reduced OhrR and 5% for OhrR-*ohrA* operator complex.

Chapter 4.

Manuscript #2

Structure of the SlyA-citrate complex: insight into the inducer binding pocket of the MarR family

Minsun Hong[1], Stephen J. Libby[2]* and Richard G. Brennan[1,3]*

[1]Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, Oregon 97239-3098, U.S.A.

[2]Department of Microbiology and Laboratory Medicine, University of Washington, Seattle, WA 98195, U.S.A.

[3]Department of Biochemistry and Molecular Biology, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, U.S.A.

Running title: Crystal structure of the SlyA-citrate complex

*Corresponding author information

[2]Department of Microbiology and Laboratory Medicine, Box University of Washington School of Medicine, Box 357110, Health Sciences Building, K-451, 1959 NE Pacific Street, Seattle, WA 98195, U.S.A.; Telephone number: 001-206-4941; Fax: 001-206-1575; E-mail: slibby@u.washington.edu

[3]Department of Biochemistry and Molecular Biology, University of Texas MD Anderson Cancer Center, Unit 1000, 1515 Holcombe Boulevard, Houston, TX 77030, U. S. A. Telephone number: 001-713-834-6390; Fax: 001-713-834-6397;

E-mail: rgbrenna@mdanderson.org

Abstract

Salmonella typhimurium SlyA is a member of the multiple antibiotic resistance regulator (MarR) family of transcription regulators. SlyA activates or represses a variety of genes that are involved in bacterial virulence, intracellular survival in macrophages, oxidative stress response and pathogenesis. The complete gene regulatory mechanism of SlyA is unclear because its inducer(s) have yet to be identified and no structures of SlyA have been determined. In order to begin to unravel the structural mechanism of SlyA, we carried out the crystal structure determination of the SlyA bound to citrate, a component of the crystallization buffer. DNA binding studies reveal that citrate binding to SlyA specifically stimulates dissociation of this DNA binding protein from the high affinity *pagC* operator. Examination of the citrate binding pocket reveals additional, contiguous space that would readily allow larger aromatic carboxylates to bind. Significantly, the location of the bound citrate overlaps the "inducer" binding site of MarR family member, OhrR, suggesting a common site responsible for the allosteric induction of all or nearly all the MarR family members.

Introduction

The *slyA* gene was first isolated from the chromosome of *Salmonella enterica* serovar Typhimurium and initially thought to be a hemolytic or cytolytic virulence factor and thus named Salmonella cytolysin by virtue of its ability to confer a hemolytic phenotype on *Escherichia coli* (*E. coli*) K12 cells (Libby et al., 1994). When introduced in *trans*, overexpression of the *slyA* gene also increases the hemolytic activity of *E. coli* K12 (Ludwig et al., 1995). However, subsequent studies revealed that SlyA is a transcription regulator of the cryptic *E. coli* chromosomal gene, *hlyE*, (also referred to *clyA*, or *sheA*), which encodes hemolycin E, a novel pore forming protein with hemolytic activity (del Castillo et al., 1997; Ludwig et al., 1999; Oscarsson et al., 1999; Wallace et al., 2000). *slyA* homologues have been identified in other *Enterobacteriaceae* including *Escherichia coli* and *Shigella flexneri* and it has been suggested that *E. coli* SlyA also functions as a hemolytic activator by binding to the *clyA* promoter region (Oscarsson et al., 1996).

Mutational studies on *slyA* indicated that this gene is required for both virulence and the intracellular survival of Salmonellae in macrophages (Libby et al., 1994). As one of its functions SlyA plays a role in the protection of these bacteria against the Reactive Oxygen Species (ROS) that are generated by the host during infection (Buchmeier et al., 1997a). Proteomic and genomic studies have shown that both *E. coli* and *Salmonella* serovar Typhimurium SlyA are involved in the transcription regulation of multiple stress response proteins and implicate functions for SlyA not only in intracellular survival but also in replication of those bacteria in phagocytic host cells (Spory et al., 2002). More recently, SlyA has been shown to be involved directly in the regulation of SsrA/B,

Salmonella Pathogenicity Island-2 (SPI-2) associated genes and collagenase, genes that are key to bacterial virulence (Fey et al., 2004; Lineham et al., 2005; Carlson, et al., 2005). SlyA has also been shown to play a role in resistance to antimicrobial peptides (Narvare et al., 2005; Shi et al., 2004).

Salmonella typhimurium (St) SlyA is a 144 amino acid residue protein (MW ~16.5 kDa) that belongs to the multiple antibiotic resistance regulator (MarR) family of transcription regulators. Members of the MarR family regulate the expression of resistance genes to multiple antibiotics, organic solvents, detergents and oxidative stress agents (Seoane and Levy, 1995; Sulavik et al., 1995; Wilkinson and Grove, 2006). Despite low primary sequence similarity amongst MarR family proteins, less than 25% on average, the members of this family show significant structural homology. x-ray crystal structures of several proteins of the MarR family, including MarR (Alekshun et al., 2001), MexR (Lim et al., 2002), OhrR (Hong et al., 2005), SarR (Liu et al., 2001) and SlyA-like protein (Wu et al., 2003), have revealed a common triangular shape with winged helix-turn-helix (wHTH) DNA-binding motifs at two of the corners of these dimeric proteins. Superimposition of structures of several MarR family proteins reveals rmsd as large as 1.9 Å and highlights their inherent conformational plasticity (Hong et al., 2005; Lim et al., 2002). Such plasticity likely facilitates the transitions between the DNA-bound, inducer-bound and apo states of each protein. Indeed superimposition of the subunits of the structures of reduced, apo OhrR and DNA bound protein, which is also reduced, reveals an rmsd of ~1.6 Å. In order to bind operator DNA, the wHTH motifs of the subunits of the reduced apo OhrR dimer rotate ~25° and the tips of the minor groove

binding wings translocate ~16 Å. Similar conformational changes are likely to occur in all MarR family members, including SlyA, to allow binding to their cognate DNA sites.

Many MarR family members bind small molecule effectors or undergo reversible chemical modifications to effect induction. For example, DNA binding by *E. coli* MarR is inhibited by salicylate (Alekshun et al., 2001) and the *B. subtilis* OhrR protein is induced by the reversible oxidation of a redox sensing residue, C15, to sulphenic acid by organic hydroperoxides (Fuangthong and Helmann, 2002). To date the complete regulatory mechanism utilized by SlyA is unclear because its inducer has yet to be identified. However, specific DNA binding by SlyA has been demonstrated and a putative consensus SlyA binding motif, TTAGCAAGCTAA, has been proposed (Stapleton et al., 2002). Consequently, *Salmonella* genome data mining and phenotypic analyses revealed that SlyA is directly or indirectly involved in the regulation of multiple genes including *fliC*, *iroN*, *pagC* and *ompC*, all of which contain one or more copies of this consensus SlyA binding site (Stapleton et al., 2002). More recently, DNase I footprinting and electrophoretic mobility shift assays (EMSA) have demonstrated a direct interaction between purified SlyA protein and the *pagC* promoter (Navarre et al., 2005).

In order to gain a fuller understanding of the mechanism of St SlyA, we set out to determine the structure of this transcription regulator in its "apo" form. Fortuitously, a buffer component, citrate, cocrystallized with the protein and hence we describe the crystal structure of SlyA in complex with citrate. Remarkably, the location of the bound citrate overlaps the redox sensing site of MarR family member OhrR (Hong et al., 2005)

and thus is consistent with a role as a SlyA inducer, albeit a poorer one or an inducer mimic (or a pseudo-inducer): a supposition that is supported by our DNA binding studies. Moreover, the size and chemical nature of the citrate binding pocket provides insight into the identity of potential high affinity, physiologically relevant inducers of SlyA.

Experimental Procedures

Preparation and Crystallization of SlyA

Recombinant SlyA was expressed and purified as previously described previously (Navarre et al., 2005). Briefly, cell lysates in 50 mM Tris, pH 7.4, 300 mM NaCl, 5 mM imidazole, 5% glycerol and 2 mM β-mercaptoethanol were applied to a column of Ni-NTA resin (Qiagen) (Figure 4. 1). Pure SlyA was eluted using an imidazole gradient ranging from 5 mM to 750 mM. Protein concentration was estimated using an extinction coefficient of 11380 $M^{-1} \bullet cm^{-1}$ at 280 nm. Selenomethionine substituted SlyA was prepared using the methionine inhibitory pathway (Doublie, 1997) and purified as described above. SlyA was crystallized at room temperature using the hanging drop vapor diffusion method. Three μl of SlyA at concentrations ranging from 7 to 10 mg/mL were mixed with 3 μl of a reservoir solution of 16% PEG 1450, 0.1 M lithium sulphate, and 0.1 M sodium phosphate/citrate, pH 4.5. The crystals completed their growth within a week (Figure 4. 2). Crystals were cryoprotected by increasing the concentration of PEG1450 incrementally to a final concentration of 35% and flash frozen directly in a liquid nitrogen stream prior to data collection.

SlyA equilibrium binding to operator DNA

The binding affinity of SlyA for the *pagC* operator was determined by fluorescence polarization experiments (Lundblad et al., 1996) utilizing a PanVera Beacon 2000 Fluorescence Polarization System (PanVera Corporation). Binding is monitored as a function of the increased polarization or anisotropy of the fluoresceinated oligodeoxynucleotide, encompassing the *pagC* operator, as SlyA is titrated into a 1 mL solution that contains 0.5 nM operator DNA. All binding experiments were carried out at 25 °C. The data are fit to the following equation to determine the binding constant (K_d):

$P = \{(P_{bound} - P_{free})[protein]\}/(K_d)+\{[protein])\}+P_{free}$

where P is the measured polarization at a given total concentration of SlyA; P_{free} is the initial polarization of the free operator; and P_{bound} is the maximum polarization of SlyA specifically bound to DNA. Binding assays were done in triplicate in 25 mM Tris, pH 7.4, 100 mM NaCl and 1 mM EDTA. Citrate inhibition assays were carried out under identical buffer conditions but with the addition of 6 mM sodium citrate pH 7.4, to the standard binding buffer. This concentration of citrate results in a ten fold reduction in binding affinity of SlyA for the *pagC* operator when compared to similar binding in the absence of citrate.

Assay for functional SlyA: Complementation of a *slyA::tet* mutant-

Strains and Plasmids

Wild type *Salmonella enterica* serovar Typhimurium 14028s or an *slyA::tet* mutant was transduced with a P22 lysate made on *pagC::lacZ* (JSG1639) (ref Prouty). Phage transductants were then transduced with a P22 lysate made on TN3740 (*leuBCD*484

trp::[Spc(r) T7 RNA polymerase *lacp lacI*]. Phage free transductants were then transformed with pSL2573 (pET16B::*slyA*) or the individual SlyA point mutants.

SlyA point mutant construction

Specific amino acid SlyA mutants were generated using a QuikChange site-directed mutagenesis kit (Stratagene). We used pSL2573, pET16B::SlyA (Navarre et al., 2005) as the template for the mutagenesis. This plasmid contains the *slyA* coding region with a 10 His-tag on the N-terminus. Confirmation of mutation construction was verified by sequence analysis. Each mutant construct was assayed for the ability to produce full-length protein. Plasmid DNA was transformed into BL21 (DE3) pLysS, induced with 0.002 M IPTG, and total protein separated by a 4-20% SDS PAGE.

Assay for complementation

Overnight cultures of *Salmonella slyA::tet* mutant, *pagC::lacZ* and expressing T7 RNA polymerase with wild type SlyA or specific point mutant plasmids were diluted 1:200, grown to an OD_{600} of 0.5. SlyA expression was then induced with 2 mM IPTG. The expression of *pagC::lacZ* was measured using standard methodology (Miller, 1972).

x-ray intensity data collection and processing and structure refinement

MAD intensity data were collected at 100 K using beamline 8.2.1 at the Advanced Light Source, Berkeley. CA. Intensity data were processed with MOSFLM and merged using SCALA. Based on the systemic absences, the space group was determined to be $P2_12_12$ with unit cell dimensions of a = 63.6 Å, b = 77.8 Å and c = 85.9 Å. Selected data

collection statistics are listed in Table 4.1.

The structure of SlyA was determined using the MAD phasing method (Hendrickson, 1991). Two of six selenium sites per asymmetric unit were located and MAD phases were calculated to 3.15 Å resolution using the SOLVE (Terwilliger and Berendzen, 1999). Phases were improved using solvent density modification and a solvent content of 56% as implemented in the CNS (Brünger et al., 1998). Subsequent electron density maps revealed, as expected, a mostly α helical structure. The selenium atoms, one site per subunit, were used as a guide to register of the protein sequence. Refinement of the initial model, which contained residues 4-75, 90-144, was carried out using CNS with one round of rigid body refinement followed by a round of simulated annealing. After model rebuilding in O (Jones et al., 1991b), rounds of positional and thermal parameter refinement were carried out followed by manual model rebuilding. Five percent of the intensity data that were chosen randomly were set aside for cross validation. Water molecules were added using the criteria of overlapping electron density whereby: F_O -F_C > 3.0 σ and 2 F_O -F_C> 1.0 σ. The final model was verified by inspection of the simulated annealing composite omit maps and F_O -F_C difference electron density map. The final model includes residue 4 - 76 and 90 - 144 of each subunit, one well ordered citrate per subunit and four water molecules per dimer. The final R_{work} and R_{free} of the MAD structure are listed in Table 1. The Ramachandran plot analysis was performed using PROCHECK (Laskowski *et al.*, 1993) and revealed 97 % of the residues lie within the allowed regions of the phi/psi plot. Seven residues are found in the disallowed region. These residues are located in the loop between beta strands β2 and β3 and have poorer

electron density. However, the removal of each or "fixing" their ϕ/φ angles to conform to favored regions results in higher R_{free} values. Thus, we chose to retain these residues in the current model.

Figures were made using Swiss-Pdb Viewer (Guex and Peitsch, 1997), POV-Ray (www.povray.org), GRASP (Nicholls et al., 1991), O and Pymol (DeLano, 2002). The coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 2AM0.

Cavity detection and calculation

Cavities were detected and their volumes were calculated by Voidoo (Kleywegt and Jones, 1994). Two dimers of the present SlyA structure without citrate were used, because there were two subunits in the asymmetric unit (ASU) of the SlyA crystal and the biologically relevant dimer is generated by one subunit and a crystallographically related dyadic mate. van der Waals Cavities were detected on a 0.5 Å grid using a probe radius of 1.4 Å. Ten volume calculation-refinement cycles were carried out using a grid-shrink factor of 0.9. Five cavities per each subunit and one at the dimerization, in total eleven cavities per dimer were detected. Volumes of cavities which coincided with the citrate binding sites were 195.8 ± 6.8 $Å^3$ and 195.6 ± 6.7 $Å^3$ in one dimer and 200.7 ± 17.1 $Å^3$ and 201 ± 9.1 $Å^3$ in the other dimer.

Results and discussion

Structure of the SlyA-citrate complex

The structure of the SlyA-citrate complex was solved by multiple anomalous wavelength dispersion (MAD) (Hendrickson, 1991) and refined to 3.15 Å resolution (Table 4. 1). The asymmetric unit (ASU) of the SlyA crystal contains two subunits. The biologically relevant dimer is generated by one subunit and a crystallographically related dyadic mate. The fold of SlyA is very similar to that of all other MarR family members and consists of six α helices and two β strands with topology: α1, (residues 5-23), α2 (residues 31-40), α3 (residues 51-54), α4 (residues 58-67), β1 (residues 73-75), β2 (residues 91-93), α5 (residues 95-114) and α6 (residues 120-137) (Figure 4. 3A). Residues 76-89 are not included in the current structure due to the weak electron density in this portion of the protein. A notable difference between SlyA and the other MarR family proteins is the structural element, which connects helices α2 and α3. In SlyA this nine residue stretch forms a loop but in the other members of the MarR family some of the corresponding residues form the β strand that precedes their helix-turn-helix DNA binding motifs and buttresses the β hairpin, which comprises part of the minor groove interacting wing. Each subunit of SlyA is divided into two functional domains: the dimerization domain, which involves the N and C termini, α1, α2, α5 and α6 and buries 5,033 Å accessible surface area, and the winged HTH (wHTH) DNA binding domain, which consists of α3, α4, β1 and β2.

Citrate is a specific *in vitro* inducer of SlyA

The structure of SlyA was determined using crystals that were grown in the presence of 100 mM sodium citrate/phosphate buffer. In the present structure, one citrate molecule per SlyA subunit is found in a crevice at the dimerization interface formed by N-terminus of α1, α1′, α2′ and α5′ (where prime indicates residues from the second subunit) (Figure 4. 3). All three carboxylates of the citrate form electrostatic interactions or hydrogen bonds with the side chains of residues Arg17′ and Arg111′ and Ile107′ as well as the main chain amide nitrogens of residues Leu5, Gly6 and Ser7 (Figure 4. 3B). In addition, the side chains from His38′ and Trp16′ make weaker hydrogen bonds to the central carboxylate group of citrate.

The binding of citrate to SlyA was unanticipated and the physiological relevance of such binding is currently unknown. One reason for citrate binding to SlyA is that it is simply an artifact and a result of the concentrations of citrate present in the crystallization buffer. Alternatively, citrate could be a small molecule activator, a true inducer or more likely an inducer mimic of SlyA, the latter of which would be a compound that shares chemical features of negative charges with the physiologically relevant inducer(s) that upon binding to SlyA lowers the affinity of the regulator for cognate DNA allosterically. The allosteric nature of an inducer or inducer mimic is different from a simple competitive inhibitor.

In order to assess the biochemical relevance of citrate binding to SlyA, this tricarboxylic acid was tested for its ability to inhibit or activate DNA binding of SlyA to the *pagC*

operator utilizing a fluorescence polarization based DNA binding assay. In the absence of citrate, SlyA binds the *pagC* operator with a K_d of ~65 nM. However, in the presence of 6 mM citrate the K_d increases to 720 nM (Figure 4. 4A). At higher citrate concentrations, the K_d is increased nearly 100 fold and even in the presence of 2 mM citrate, which is ~5-6 fold higher than the concentrations of intracellular citrate found in the closely related Gram negative bacterium *E. coli* (Varghese et al., 2003), the DNA binding affinity is diminished up to 4 fold (M.H. & R.G.B, data not shown).

To ensure that this citrate effect on SlyA-*pagC* operator binding was specific and not a general property of the negative charge of this molecule, which might allow it to act as a simple competitor of the DNA phosphate backbone, the identical binding experiments were performed with the *Bacillus subtilis* DNA binding proteins OhrR and BmrR. OhrR is a member of the MarR family in which a unique cysteine residue, C15, can be oxidized by organic hydroperoxides, which leads to the derepression of *ohrA*, the gene encoding an organic hydroperoxidase (Fuangthong et al., 2002). BmrR is a transcription regulator that belongs to the MerR family (Brown et al., 2003; Summers, 1992). Interestingly, BmrR bound to the *bmr* operator has been crystallized in the absence and presence of similar citrate concentrations (100 mM citrate) (Heldwein and Brennan, 2001). Citrate was not found in the wild type protein under conditions in which citrate was present. In accord with the hypothesis that the inhibitory effect of citrate binding to SlyA is specific, the presence of 6 mM citrate caused no difference in the ability of OhrR or BmrR to bind their cognate DNA sites as their K_ds remain virtually identical (Figures 4. 4B and C). Higher concentrations of citrate were equally ineffective (data not shown). Therefore,

citrate is not a simple nonspecific inhibitor of DNA binding by SlyA, but rather a specific inducer or inducer mimic.

In order to gain additional insight into the chemical nature of other biologically relevant inducers or at least those that might bind with higher affinity, an analysis of all cavities of SlyA was carried out using Voidoo (Kleywegt and Jones, 1994). After removal of the bound citrates, one per subunit, eleven cavities per SlyA dimer (five cavities per each subunit and one at dimerization interface) were identified. Two coincided with the citrate binding site and the sum of their contiguous volumes combined was ~200 $Å^3$ (Figure 4. 5A). Residues that surround these cavities include Gly6, Ser7, Ala10, Arg14, Arg17′, Trp34′, His 38′, Asn39′ and Gln42′ (truncated to alanine). The citrate binding site is located towards the surface of the SlyA, at the one end of cavity leaving a large unfilled volume (Figure 4. 5B). Thus, a higher affinity SlyA inducer likely utilizes more or the entire binding pocket that is only partially occupied by citrate. Beyond its polar nature, the presence of aromatic residues in the pocket, Trp34′ and His38′, and basic residues, Arg14 and Arg17′, which can be involved in stacking interactions with π electrons, suggests that high affinity and likely biologically relevant inducers of SlyA have in addition to one or more carboxylates, an aromatic component.

Comparison to MarR family members

The superimposition of each independent SlyA subunit onto those of MarR, OhrR and SlyA-like protein results in rmsds of 1.8 Å, 2.0 Å and 1.8 Å, respectively, underscoring the high structural similarity but inherent conformational plasticity of MarR family

proteins (Lim et al., 2002; Hong et al., 2005) (Figure 4. 6). As expected, the superimposition of the structures of SlyA and OhrR in its DNA bound form, which is the only MarR family to have been crystallized bound to cognate DNA (Hong et al., 2005), revealed conformations that are incompatible with DNA binding. In the OhrR-DNA complex, protein-DNA interactions are made from the Helix-Helix elements (helices α1 and α2) to the DNA backbone proximal to the operator dyad, the Helix-Turn-Helix motifs to consecutive major grooves and the Winged regions to the minor grooves and backbone of the operator DNA. More specifically, the distance between the loops connecting Helix-Helix motifs of DNA bound OhrR are ~6 Å longer in SlyA, as measured from the Cα of Gln31 to the Cα of Gln31′ (~19 Å) for SlyA and from the corresponding Cα atoms for OhrR (Thr39-Thr39′, ~25 Å). In addition the distance between corresponding Cα atoms located at the N-termini of recognition helices, α4 and α4′ is enlarged by 3 Å (~26 Å) in the SlyA structure. These conformational features preclude the SlyA-citrate complex from binding to the DNA structure as taken in OhrR-operator DNA complex.

In addition to the HH and HTH motifs, the Wings of all MarR family members, which are quite extended, are assumed to interact with distal upstream and downstream minor grooves (Hong et al., 2005). In the citrate bound form of SlyA the tip of each wing is disordered due to their expected conformational flexibility in the absence of cognate DNA. However, a significant role of the Wing in DNA binding affinity had been noted in previous mutational studies on MarR (Alekshun et al., 2001) and MexR (Saito et al., 2003). The recent crystal structure of the OhrR-*ohrA* operator complex revealed that

Wing residues Asp92 and Arg94, which are highly conserved amongst the MarR family (Figure 4. 3C), form an interaction network to contact the minor groove in which Asp92 buttresses the position of Arg94, the side chain of which makes a contact with a pyrimidine base (Hong et al., 2005). To ascertain whether or not SlyA uses a similar interaction network, homologous residues Asp84 and Arg86 were substituted with alanine individually and their ability to bind the *pagC* promoter tested. Such changes results in SlyA proteins that bind this cognate DNA site with equilibrium dissociation constants of 1,127 ± 105 nM for Asp84Ala and 1,112 ± 132 nM for Arg86Ala, values that are ~15 fold higher than wild type SlyA (Figures 4. 4A and D). Moreover, *in vivo* complementation studies reveal that unlike the wild type protein, SlyA mutants Asp84Ala, Arg86Ala and the double mutant Asp84Ala/Arg86Ala cannot activate transcription of the *lacZ* gene from the wild type *pagC* promoter indicating their inability to bind a cognate DNA sequence (Table 4. 2).

A common inducer binding pocket for the MarR family?

Clearly the folds of all MarR family members will be similar despite their observed conformational plasticities. This begs the question will all members of the MarR family contain and use a similarly located inducer binding pocket to elicit induction? In the well characterized LacI/GalR family of bacterial gene regulators a commonly located small molecule effector binding pocket has been described structurally for members PurR (Schumacher et al., 1994; Schumacher et al., 1997), LacI (Lewis et al., 1996; Bell and Lewis, 2001) and CcpA (Schumacher and Brennan, unpublished data). In order to investigate an analogous possibility for the MarR family, the electrostatic surface

potentials were calculated for SlyA, OhrR, MarR and MexR (Figure 4. 7). Strikingly, the citrate binding pocket of SlyA and the reactive site of OhrR, which contains the redox sensor residue Cys15, are located nearly identically and primarily composed of residues from helices α1, α1' and α2' (Figure 4. 7A and B). The electrostatic potential surfaces highlight the complementary chemical nature of each binding pocket for its ligand: the electropositive area of SlyA engulfing the citrate and the hydrophobic/neutral patch of OhrR surrounding residue Cys15 that makes an excellent "landing pad" for organic hydroperoxides, the compounds responsible for oxidizing the thiol group of residue Cys15 to sulphenic acid (Fuangthong et al., 2002). Although the full induction mechanism of neither protein is known, oxidation of OhrR residue Cys15 would result in steric clash between the sulphenic acid moiety and the side chains of nearby tyrosines or methionines. Any subsequent rotations necessary to alleviate such clash would cause direct changes in the conformations of the Helix-Helix elements and allosteric displacements of the wHTH motifs (Hong et al., 2005).

By contrast, MarR does not appear to utilize this common inducer binding pocket as the structure of a MarR-salicylate complex reveals two molecules of salicylate per subunit bound at sites termed SAL-A and SAL-B (Alekshun et al., 2001). At near millimolar concentrations salicylate inhibits MarR activity (Alekshun and Levy, 1999) and is used routinely as a model inhibitor of MarR in order to induce MarA expression thereby conferring the *mar* phenotype in *E. coli* and *S. typhimurium* (Cohen et al., 1993; Sulavik et al., 1995, 1997). The SAL-A site is located in a crevice between the "recognition helix" of the HTH motif and the β sheet of the wing. The SAL-B site is located on the

opposite side of helix α4 (Figure 4. 7C). When the structures of the MarR-salicylate and OhrR-*ohrA* operator complexes are overlaid the salicylates of the Sal-A site overlap with phosphate backbone of the DNA, i.e., they clash sterically with each other (data not shown). Such an observation suggests that salicylate competitively inhibits MarR binding to cognate DNA, i.e., salicylate bound to the SAL-A site is not an allosteric inducer. If true, this would imply that the MarR family has at least two, small molecule induced transcription activation mechanisms. Of course, the possibility that the salicylate binding sites in this particular MarR structure are artifacts cannot be precluded as rather high concentrations of salicylate (250 mM) were necessary for crystallization. Notably, inspection of the electrostatic potential surface of MarR reveals a strongly positively charged pocket that is distal from the SAL-A and SAL-B sites but overlapping with the respective inducer binding pocket and reactive site of SlyA and OhrR (Figures 4. 7A-C). Perhaps this pocket is utilized by physiologically germane inducers of MarR to effect an allosteric induction mechanism.

The structure of MexR, the *Pseudomonas aeruginosa* MarR family member that regulates the expression of the *mexAB* multidrug efflux pump operon (Poole et al., 1996; Srikumar et al., 2000) also supports the idea of a common allosteric inducer binding site. In that structure four independent MexR dimers were found in the asymmetric unit (Lim et al., 2002). Intriguingly, the C-terminus of helix α6 from one subunit is inserted into a surface cleft of another dimer (Figure 4. 7D). The authors speculated that this latter dimer represented the induced conformation as its structure deviated significantly from the other three dimers. Whether this assumption is correct requires the structure of a

relevant MexR-inducer complex. Interestingly, no small molecule inducer has been yet identified for MexR and the possibility that the activity of MexR is regulated by chemical modification or the binding of another protein or peptide, as perhaps mimicked in the crystal structure, remains. An analogous regulatory system has been well characterized for the LacI/GalR family member CcpA, the DNA binding of which is regulated by the binding of either HPr or Crh, only after these latter proteins have been phosphorylated on residue Ser46 by HPr kinase/phosphorylase (Deutscher et al., 2005; Gosseringer et al., 1997; Jones et al., 1997; Schumacher et al., 2004).

In conclusion our biochemical studies demonstrate that moderate concentrations of citrate specifically reduce the binding affinity of SlyA for *pagC* operator DNA and can be considered an inducer of SlyA, albeit a weaker one, or an *in vitro* inducer, which contains key chemical components of the physiologically relevant inducers of SlyA. Thus, the structure of the SlyA-citrate complex provides a view of the induced form of this transcription regulator and further inspection of the extended citrate binding pocket indicates that larger, most likely aromatic carboxylate containing compounds are biologically relevant true inducers with higher affinity. Moreover, the structure of the SlyA-citrate complex reveals an inducer binding pocket that overlaps the reactive site of OhrR and therefore implicates this location as a common site responsible for the allosteric induction of all or nearly all the MarR family members.

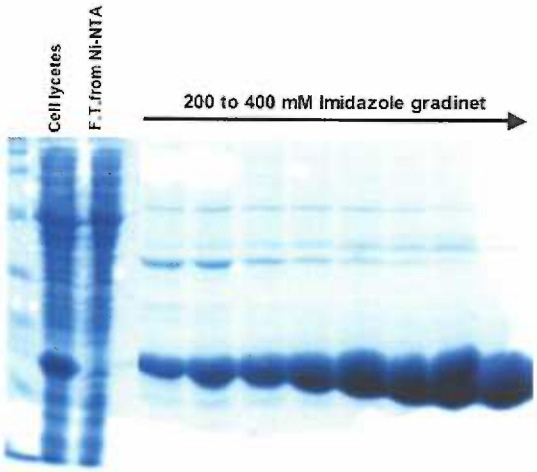


Figure 4. 1 Purification of recombinant *S. typhimurium* SlyA protein. Cell lysates of 6xHis-*S. typhimurium* SlyA protein overexpressed in *E. coli* were applied onto a Ni-NTA column. Pure SlyA was eluted by an imidazole gradient from 5 to 750mM.

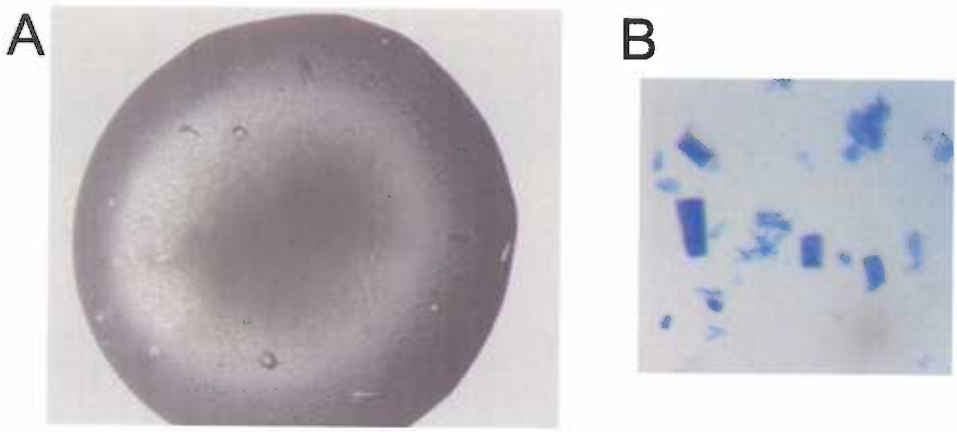


Figure 4. 2 Pictures of the SlyA-Citrate complex crystals. **B**, SlyA crystals stained with protein dye.

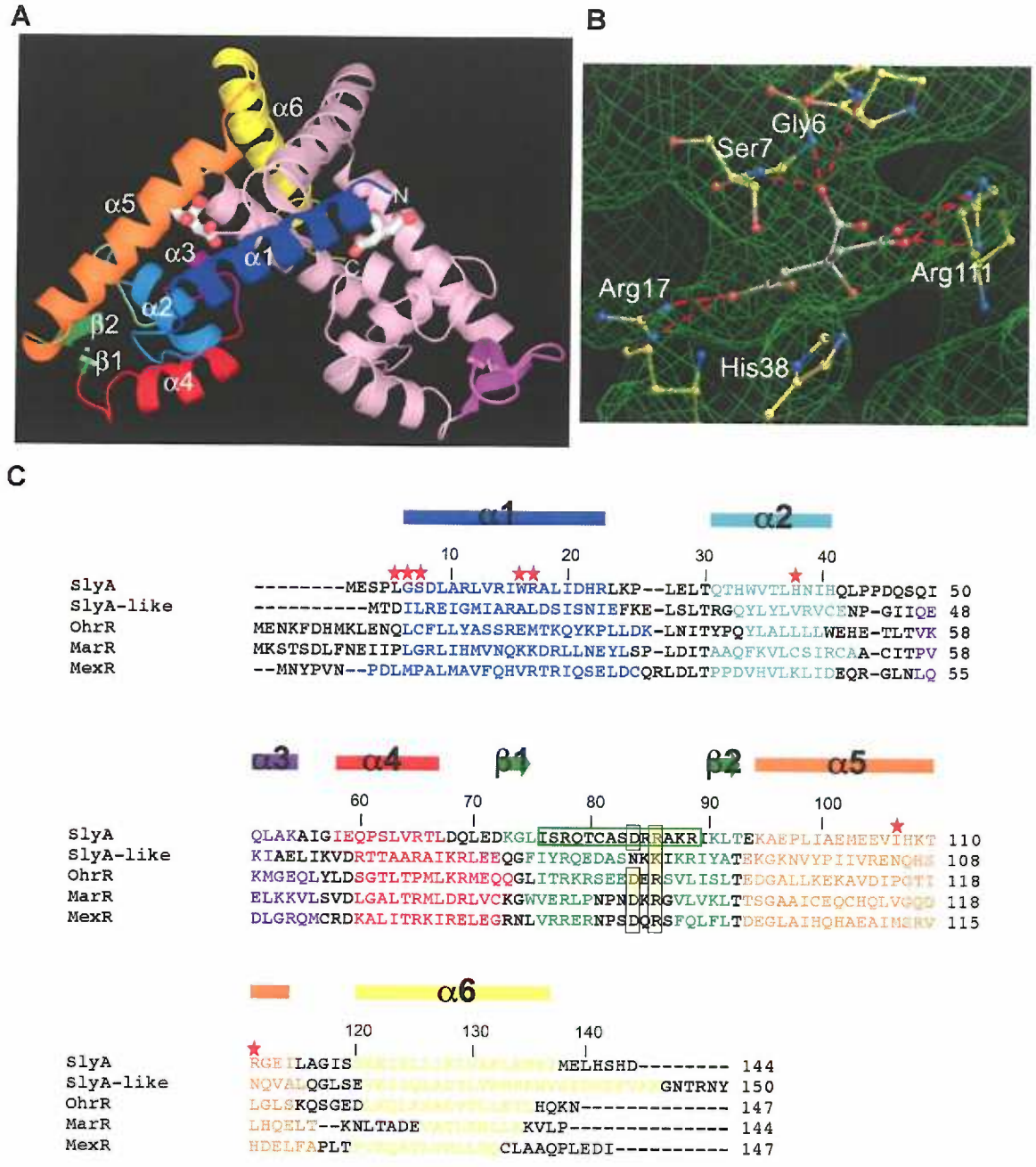


Figure 4. 3 . Structure of the SlyA-citrate complex. **A**, SlyA is shown as a ribbon diagram with the N and C termini and secondary structural elements of one subunit colored and labelled. The dyadic mate is colored in pink. Citrate molecules are shown as sticks with carbon and oxygen atoms colored grey and red, respectively. **B**, Electron

density map of the SlyA-citrate complex. The map was calculated after density modification to 3.15 Å resolution and contoured at 1σ. The refined positions of the citrate molecule and neighboring residues are shown in sticks and a prime indicates a residue from the other subunit. Red dashes indicate interactions between selected protein and citrate atoms. **C**, Structural alignment of the primary sequences of selected MarR family proteins. The secondary structure elements of SlyA-citrate are indicated above the sequence by arrows (β strands) and solid rectangles (α helices). The colour of each secondary structure matches those of Figure A. Citrate binding residues are marked with red stars. The region not shown in current SlyA structure due to its poor electron density, between β1 and β2, is boxed in green. Residues Asp84 and Arg86, located at the tip of wing in other MarR structures and mutated to alanine in this study, are shaded in yellow.

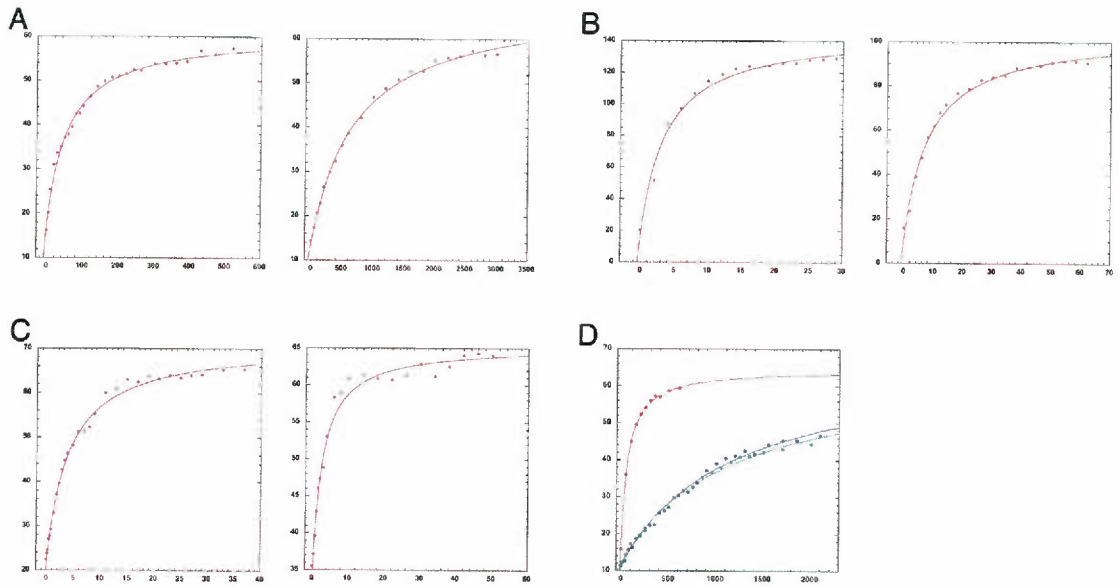


Figure 4. 4 Specific inhibition of SlyA-DNA binding by citrate. The plots of the binding isotherms of three DNA binding proteins and their high affinity operator DNA in the absence (left panel) or presence (right panel) of 6 mM citrate are shown. Millipolarization (mP) is plotted against the protein concentration. **A**, SlyA-*pagC* operator binding. **B**, OhrR-*ohrA* operator binding. **C**, BmrR-*bmr* operator binding. **D**, SlyA mutants D84A (blue plot), R86A (green plot) and wild type SlyA (red plot) binding to the *pagC* operator.

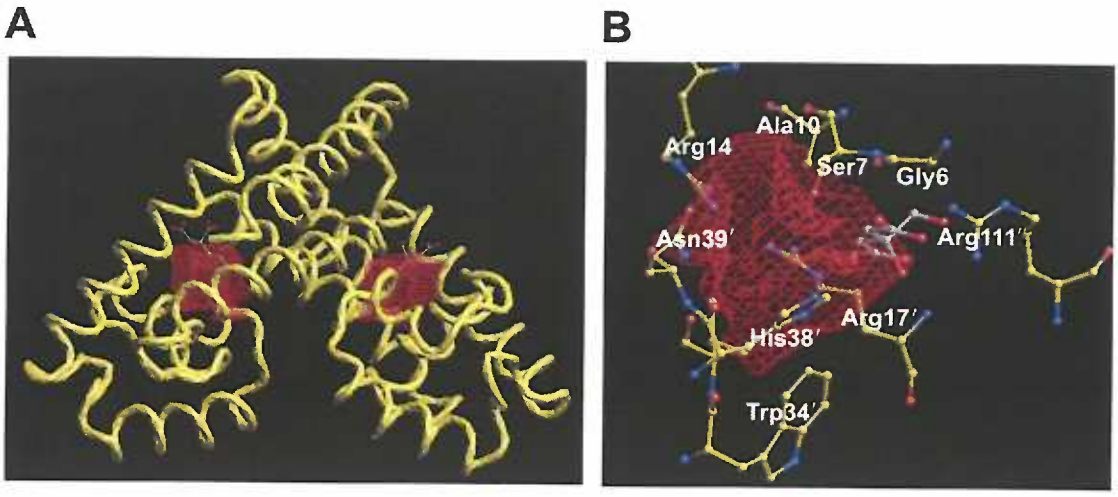


Figure 4. 5 The inducer binding pocket of SlyA. **A**, The SlyA dimer is shown as a yellow ribbon and the cavity that forms the inducer binding site, as revealed by Voidoo (Kleywegt and Jones, 1994) is shown as a red mesh. The bound citrate is shown as a ball and stick model in which oxygens are coloured red and carbons, grey. **B**, Close up of the inducer binding cavity that is contiguous with the citrate binding site. Residues encasing the cavity are shown as balls and sticks and colored according to atom type. Primes indicate residues from the other subunit.

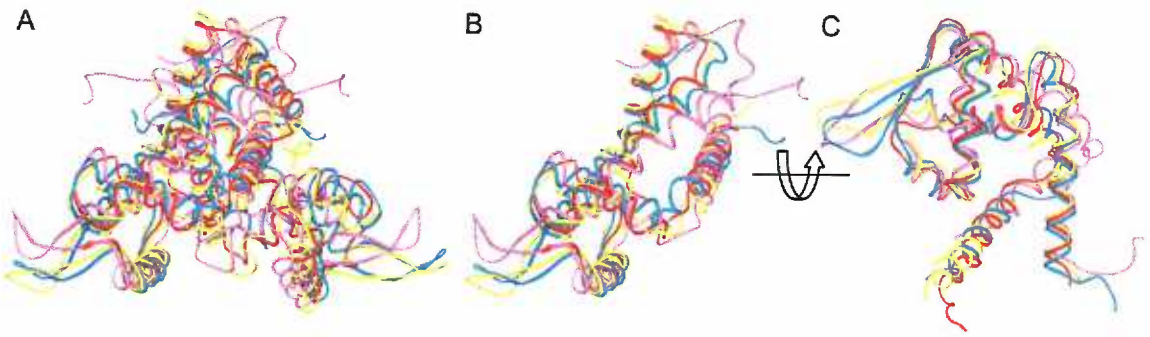


Figure 4. 6 Conformational plasticity of the MarR family. Overlays of the structures of SlyA (PDB ID code 2AM0), OhrR-DNA (1Z9C), MarR (1JGS) and SlyA-like (1LJ9) dimers (**A**) and monomers (**B** and **C**), colored in red, blue, pink and yellow, respectively. Monomers were superimposed using the main chain atoms. **C**, The overlays rotated by 90°.

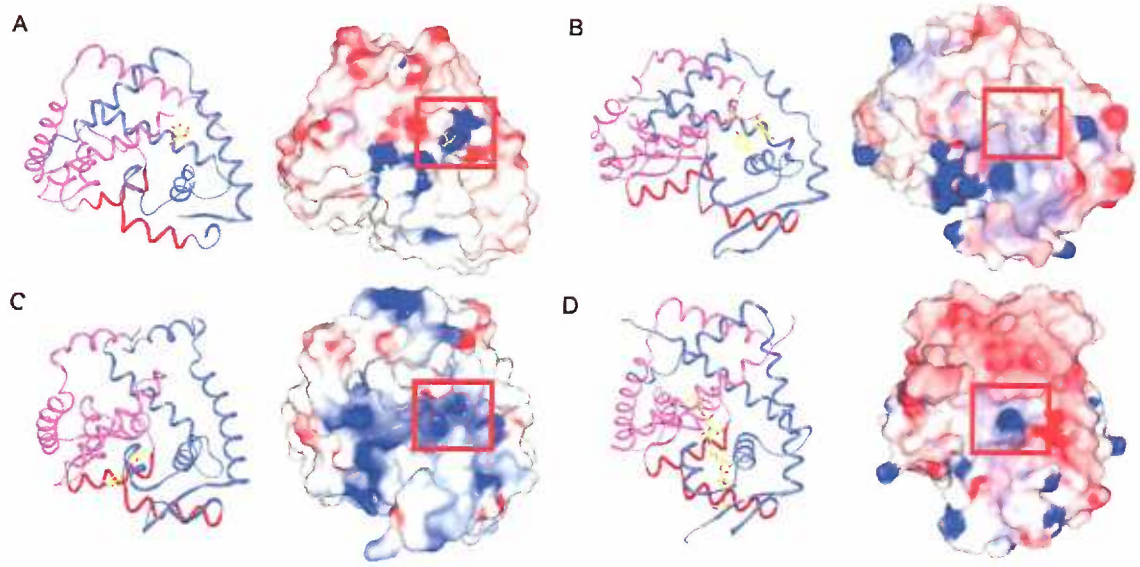


Figure 4. 7 A common inducer binding site of the MarR family. The electrostatic surface potentials of SlyA, OhrR, MarR and MexR were calculated using GRASP (Nicholls et al., 1991) and their structures superimposed and displayed in the same orientation. The putative inducers or inhibitors of SlyA, MexR and MarR are show as sticks. The common location of the known or surmised inducer binding pockets of each protein is boxed in red. **A**, The SlyA-citrate complex. **B**, Reduced OhrR, in which the redox sensor residue Cys15 and Tyr29′ and Tyr40′ are shown as sticks. **C**, The MarR-salicylate complex. Note the boxed positively charged patch that is located similarly to SlyA and OhrR inducer binding sites. **D**, MexR bound to the α helix of another subunit in the asymmetric unit. The major groove binding recognition helices of each subunit of each protein, α4, are coloured red to highlight the distal location of the inducer binding pockets.

Table 4. 1 Selected crystallographic data and statistics

Data Collection and Phasing			
Wavelength (λ)	0.9794	0.9184	0.9796
Resolution (Å)		77.82-3.15	
Overall R_{sym} [a]	0.04 (0.21)[b]	0.07 (0.22)	0.06 (0.19)
Overall I/σ(I)	9.3 (3.7)	6.3 (3.4)	8.9 (4.0)
Total Reflections (#)	26,859	26,717	26,896
Unique Reflections (#)	7,545	7,542	7,548
Completeness (%)	99.8 (100)	99.8 (100)	99.8 (100)
Overall Figure of Merit[c]		0.48 (0.86)	
Refinement Statistics			
R_{work}/R_{free}(%)[d]		25.5/29.7	
Bond angles (°)		0.008	
Bond lengths (Å)		1.37	
B values (Å^2)		90.8	
Ramachandran analysis			
Most favoured (%/#)		79.7 (181)	
Add. allowed (%/#)		14.1 (32)	
Gen. allowed (%/#)		3.1 (7)	
Disallowed (%/#)		3.1 (7)	

[a] $R_{sym} = \Sigma\Sigma|I_{hkl}-I_{hkl(j)}|/ \Sigma I_{hkl}$, where $I_{hkl(j)}$ is the observed intensity and I_{hkl} is the final average intensity value.

[b] values in parentheses are for the highest resolution shell.

[c] Figure of Merit = $<|\Sigma P(\alpha)e^{i\alpha}/\Sigma P(\alpha)|>$, where α is the phase and $P(\alpha)$ is the phase probability distribution.

[d] $R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$ and $R_{free} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$; where all reflections belong to a test set of 5% randomly selected reflections.

Table 4. 2 Complementation of a *slyA::tet* mutant

Mutation	Complementation[a]
Wild type	+
Asp84Ala	-
Arg86Ala	-
Asp84A/Arg86A	-

[a]Complementation was determined qualitatively by measuring *pagC::lacZ* expression. A plus sign indicates SlyA proteins that were able to activate expression of the *lacZ* gene, which results in blue colonies. A negative sign indicates SlyA proteins that were unable to activate expression of the *lacZ* gene, which results in white colonies.

Chapter 5. Summary and conclusions

In this thesis, X-ray crystal structural studies on the MarR family proteins, OhrR-DNA and SlyA-pseudoinducer complexes were undertaken to provide an atomic description and understanding of the mechanism of transcription regulation of ROS response by OhrR and to gain general insight into MarR family transcription regulation.

The X-ray structures of OhrR in its reduced (CysSH) and DNA bound (OhrR- *ohrA* operator DNA complex) forms have been determined at 2.5Å and 2.64 Å resolution, respectively. The OhrR-*ohrA* operator complex provides the first atomic description of the binding and recognition of DNA bound MarR-family member. The structure reveals that the operator DNA by OhrR are made by the Helix-Helix motif and Winged-Helix-Turn-Helix motif and the Helix-Helix motif. The Helix-Helix motif is a novel DNA binding element, the Helix-Helix motif and the Winged-Helix-Turn-Helix motif of OhrR is a chimera, in which the HTH motif is clearly bacterial and the minor groove binding Wing is clearly eukaryotic. Thus, the MarR family is a structural intermediate between the gene regulators of prokaryotes and eukaryotes. We propose that such DNA binding modes will be used throughout MarR family.

In addition, comparisons of the structures of the apo reduced form and DNA bound OhrR, SlyA-citrate and other MarR family proteins reveal the conformational plasticity of this family of relatively small proteins. Such plasticity is likely critical to regulation in MarR proteins for DNA binding and induction or activation mechanisms by their effector molecules.

Moreover, the structure of OhrR is the first of the MarR family, in which the reactive site or inducer binding site is truly defined. The location of the OhrR reactive site, between the two subunits of the dimer, is in our view the most probable area of the ligand binding sites of nearly all, if not all, MarR family members. The induction of OhrR requires the oxidation of the side chain of Cys15 to sulphenic acid. This necessitates the nucleophilic thiolate form of the side chain of this peroxidant sensor and hence, an ionisable sulphydryl group at neutral pH. Unlike the well studied redox regulator, OxyR, which uses nearby basic residues to drive down the pK_a of the sulphydryl group of its sensor cysteine, there are no basic residues within 7 Å of the sulphur atom of Cys15. Based on the structural observation we proposed that OhrR lowers pK_a by placing Cys15 at the N-terminus of helix α1 and engaging its SH moiety in hydrogen bonds with the phenolic oxygens of two tyrosines on the other subunit. Supporting our hypothesis, pK_a of Cys15 was determined to substantially low ~6. 4. Hence, the structure of OhrR has revealed a novel mechanism of pK_a modulation by a gene regulator. Thus, the N-terminally located C15 is in the thiolate form at physiological pH and should be ready to react with OHPs.

In addition to their role in stabilizing the reactive site cysteine side chain, residues Y29 and Y40 enforce a steric constraint upon the OhrR reactive site such that oxidation of the C15 side chain to sulphenic acid would result in steric clash with their phenolic side chains if no compensatory conformational change occurred. Such observation suggests that any conformational change in the reactive site (or the effector molecule binding site) induces further changes of OhrR as well as MarR family protein structures to make

operator DNA binding disfavored. This ligand induction mechanism can be further supported by the inherent structurally flexible nature of MarR family members.

Finally the structure of the SlyA-citrate complex reveals an inducer binding pocket that overlaps the reactive site of OhrR and therefore implicates this location as a common site responsible for the allosteric induction of all or nearly all the MarR family members. Superimposition of structures Sly-citrate and OhrR-DNA complexes indicates that the inducer binding the DNA binding motif into an improper conformer of on DNA molecules. The citrate binding pocket in SlyA was observed between the two subunits of the dimer as same position as the reactive site of OhrR where the reactive cysteine, residue 15, is located. Even it is not clear the physiological role of SlyA induced by citrate molecules *in vivo* we were able to establish that the citrate molecule specifically induces the depression of SlyA on operator DNA dissociation. Thus we expand our observation in inducer binding pocket in the structure of SlyA-citrate complex and the reactive site in that of OhrR to the ligand binding sites nearly all, if not all, MarR family members and their induction mechanisms.

References

Alekshun, M. N., Kim, Y. S., and Levy, S. B. (2000). Mutational analysis of MarR, the negative regulator of marRAB expression in Escherichia coli, suggests the presence of two regions required for DNA binding. Mol Microbiol *35*, 1394-1404.

Alekshun, M. N., and Levy, S. B. (1997). Regulation of chromosomally mediated multiple antibiotic resistance: the mar regulon. Antimicrob Agents Chemother *41*, 2067-2075.

Alekshun, M. N., and Levy, S. B. (1999). The mar regulon: multiple resistance to antibiotics and other toxic chemicals. Trends Microbiol *7*, 410-413.

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.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., and Matthews, B. W. (1981). Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. Nature *290*, 754-758.

Aslund, F., Zheng, M., Beckwith, J., and Storz, G. (1999). Regulation of the OxyR transcription factor by hydrogen peroxide and the cellular thiol-disulfide status. Proc Natl Acad Sci U S A *96*, 6161-6165.

Atichartpongkul, S., Loprasert, S., Vattanaviboon, P., Whangsuk, W., Helmann, J. D., and Mongkolsuk, S. (2001). Bacterial Ohr and OsmC paralogues define two protein families with distinct functions and patterns of expression. Microbiology *147*, 1775-1782.

Babior, B. M. (1984). Oxidants from phagocytes: agents of defense and destruction. Blood *64*, 959-966.

Barbosa, T. M., and Levy, S. B. (2000). Differential expression of over 60 chromosomal genes in Escherichia coli by constitutive expression of MarA. J Bacteriol *182*, 3467-3474.

Blow, D. (2002a). Molecular replacement. In Outline of Crystllography for Biologists (Oxford), pp. 162-176.

Blow, D. (2002b). Outline of Crystallography for Biologists, Oxford university press).

Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. Cell *74*, 773-776.

Brown, N. L., Stoyanov, J. V., Kidd, S. P., and Hobman, J. L. (2003). The MerR family of transcriptional regulators. FEMS Microbiol Rev *27*, 145-163.

Brünger, 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., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54 (Pt 5)*, 905-921.

Bsat, N., Herbig, A., Casillas-Martinez, L., Setlow, P., and Helmann, J. D. (1998). Bacillus subtilis contains multiple Fur homologues: identification of the iron uptake (Fur) and peroxide regulon (PerR) repressors. Mol Microbiol *29*, 189-198.

Buchmeier, N., Bossie, S., Chen, C. Y., Fang, F. C., Guiney, D. G., and Libby, S. J. (1997a). SlyA, a transcriptional regulator of Salmonella typhimurium, is required for resistance to oxidative stress and is expressed in the intracellular environment of macrophages. Infect Immun *65*, 3725-3730.

Buchmeier, N., Bossie, S., Chen, C. Y., Fang, F. C., Guiney, D. G., and Libby, S. J. (1997b). SlyA, a transcriptional regulator of Salmonella typhimurium, is required for resistance to oxidative stress and is expressed in the intracellular environment of macrophages. Infect Immun *65*, 3725-3730.

Cadenas, E. (1989). Biochemistry of oxygen toxicity. Annu Rev Biochem *58*, 79-110.

Cambell, R. (2005). Protein crystallography lecture notes. http://adeliebiochemqueensuca/~rlc/work/teaching/.

Carlsson, J., Drevin, H., and Axen, R. (1978). Protein thiolation and reversible protein-protein conjugation. N-Succinimidyl 3-(2-pyridyldithio)propionate, a new heterobifunctional reagent. Biochem J *173*, 723-737.

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

Choi, H., Kim, S., Mukhopadhyay, P., Cho, S., Woo, J., Storz, G., and Ryu, S. (2001). Structural basis of the redox switch in the OxyR transcription factor. Cell *105*, 103-113.

Christman, M. F., Storz, G., and Ames, B. N. (1989). OxyR, a positive regulator of hydrogen peroxide-inducible genes in Escherichia coli and Salmonella typhimurium, is homologous to a family of bacterial regulatory proteins. Proc Natl Acad Sci U S A *86*, 3484-3488.

Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature *364*, 412-420.

Cohen, S. P., Levy, S. B., Foulds, J., and Rosner, J. L. (1993). Salicylate induction of antibiotic resistance in Escherichia coli: activation of the mar operon and a mar-independent pathway. J Bacteriol *175*, 7856-7862.

Cook, W. J., Kar, S. R., Taylor, K. B., and Hall, L. M. (1998). Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins. J Mol Biol *275*, 337-346.

Cowtan, K. (1998). Modified phased translation functions and their application to molecular-fragment location. Acta Crystallogr D Biol Crystallogr *54 (Pt 5)*, 750-756.

Cussiol, J. R., Alves, S. V., de Oliveira, M. A., and Netto, L. E. (2003). Organic hydroperoxide resistance gene encodes a thiol-dependent peroxidase. J Biol Chem *278*, 11570-11578.

Davies, K. J. (1987). Protein damage and degradation by oxygen radicals. I. general aspects. J Biol Chem *262*, 9895-9901.

Davies, K. J., Delsignore, M. E., and Lin, S. W. (1987). Protein damage and degradation by oxygen radicals. II. Modification of amino acids. J Biol Chem *262*, 9902-9907.

del Castillo, F. J., Leal, S. C., Moreno, F., and del Castillo, I. (1997). The Escherichia coli K-12 sheA gene encodes a 34-kDa secreted haemolysin. Mol Microbiol *25*, 107-115.

DeLano, W. L. (2002). he PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, California).

Demple, B. (1998). A bridge to control. Science *279*, 1655-1656.

Deutscher, J., Herro, R., Bourand, A., Mijakovic, I., and Poncet, S. (2005). P-Ser-HPr--a link between carbon metabolism and the virulence of some pathogenic bacteria. Biochim Biophys Acta *1754*, 118-125.

Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. Methods Enzymol *276*, 523-530.

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

Engelmann, S., Lindner, C., and Hecker, M. (1995). Cloning, nucleotide sequence and reulation of katE encoding a sigma B-dependent catalse in bavillus subtilis. J Bacteriol *177*, 5598-5605.

Evans, K., Adewoye, L., and Poole, K. (2001). MexR repressor of the mexAB-oprM multidrug efflux operon of Pseudomonas aeruginosa: identification of MexR binding sites in the mexA-mexR intergenic region. J Bacteriol *183*, 807-812.

Evans, P. (1997). Scala. Joint CCP4 and ESF-EACBM Newsletter *33*, 22-24.

Fuangthong, M., Atichartpongkul, S., Mongkolsuk, S., and Helmann, J. D. (2001). OhrR is a repressor of ohrA, a key organic hydroperoxide resistance determinant in Bacillus subtilis. J Bacteriol *183*, 4134-4141.

Fuangthong, M., and Helmann, J. D. (2002). The OhrR repressor senses organic hydroperoxides by reversible formation of a cysteine-sulfenic acid derivative. Proc Natl Acad Sci U S A *99*, 6690-6695.

Fuangthong, M., Herbig, A. F., Bsat, N., and Helmann, J. D. (2002). Regulation of the Bacillus subtilis fur and perR genes by PerR: not all members of the PerR regulon are peroxide inducible. J Bacteriol *184*, 3276-3286.

Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B., and Burley, S. K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature *403*, 916-921.

Garcia-Castellanos, R., Mallorqui-Fernandez, G., Marrero, A., Potempa, J., Coll, M., and Gomis-Ruth, F. X. (2004). On the transcriptional regulation of methicillin resistance: MecI repressor in complex with its operator. J Biol Chem *279*, 17888-17896.

Giacovazzo, C., Monaco, H., Viterbo, D., and Scordari, F. (1992). Fundamentals of Crystallography (New York, Oxford University Press).

Gosseringer, R., Kuster, E., Galinier, A., Deutscher, J., and Hillen, W. (1997). Cooperative and non-cooperative DNA binding modes of catabolite control protein CcpA from Bacillus megaterium result from sensing two different signals. J Mol Biol *266*, 665-676.

Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714-2723.

Guss, J. M., Merritt, E. A., Phizackerley, R. P., Hedman, B., Murata, M., Hodgson, K. O., and Freeman, H. C. (1988). Phase determination by multiple-wavelength x-ray diffraction: crystal structure of a basic "blue" copper protein from cucumbers. Science *241*, 806-811.

Gutteridge, J. M., and Halliwell, B. (1989). Iron toxicity and oxygen radicals. Baillieres Clin Haematol *2*, 195-256.

Haldenwang, W. G. (1995). The sigma factors of bacillus subtilis. Microbiol Rev, 1-30.

Halliwell, B., and Gutteridge, J. M. (1989). Free radicals in biology and medicine, third edition (Oxford, Oxford science publications).

Hancock, R. E. (1998). Resistance mechanisms in Pseudomonas aeruginosa and other nonfermentative gram-negative bacteria. Clin Infect Dis *27 Suppl 1*, S93-99.

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.

Helmann, J. D., Wu, M. F., Gaballa, A., Kobel, P. A., Morshedi, M. M., Fawcett, P., and Paddon, C. (2003). The global transcriptional response of Bacillus subtilis to peroxide stress is coordinated by three transcription factors. J Bacteriol *185*, 243-253.

Hendrickson, W. A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science *254*, 51-58.

Hendrickson, W. A., Smith, J. L., Phizackerley, R. P., and Merritt, E. A. (1988). Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. Proteins *4*, 77-88.

Hengge-aronis, R. (2000). The general stress response in Escherichia coli. In Bacterial stress responses, G. Storz, and R. Hengge-Aronis, eds. (Washington, D. C., ASM press), pp. 161-178.

Herbig, A., and Helmann, J. D. (2002). *Bacillus subtilis* and its closest relatives, 2nd ed. edn (Washington, D.C., ASM Press).

Herbig, A. F., and Helmann, J. D. (2001). Roles of metal ions and hydrogen peroxide in modulating the interaction of the Bacillus subtilis PerR peroxide regulon repressor with operator DNA. Mol Microbiol *41*, 849-859.

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

Hol, W. G., van Duijnen, P. T., and Berendsen, H. J. (1978). The alpha-helix dipole and the properties of proteins. Nature *273*, 443-446.

Hong, M., Fuangthong, M., Helmann, J. D., and Brennan, R. G. (2005). Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family. Mol Cell *20*, 131-141.

Jones, B. E., Dossonnet, V., Kuster, E., Hillen, W., Deutscher, J., and Klevit, R. E. (1997). Binding of the catabolite repressor protein CcpA to its DNA target is regulated by phosphorylation of its corepressor HPr. J Biol Chem *272*, 26530-26535.

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

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard (1991b). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A *47 (Pt 2)*, 110-119.

Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. Acta Crystallogr D Biol Crystallogr *55 (Pt 2)*, 484-491.

Kleywegt, G. (2001). Model validation. http://xraybmcuuse/gerard/embo2001/modval/.

Kleywegt, G. J., and Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystallogr D Biol Crystallogr *50*, 178-185.

Klomsiri, C., Panmanee, W., Dharmsthiti, S., Vattanaviboon, P., and Mongkolsuk, S. (2005). Novel roles of ohrR-ohr in Xanthomonas sensing, metabolism, and physiological adaptive response to lipid hydroperoxide. J Bacteriol *187*, 3277-3281.

Kortemme, T., and Creighton, T. E. (1995). Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family. J Mol Biol *253*, 799-812.

Kullik, I., Stevens, J., Toledano, M. B., and Storz, G. (1995). Mutational analysis of the redox-sensitive transcriptional regulator OxyR: regions important for DNA binding and multimerization. J Bacteriol *177*, 1285-1291.

Lange, R., and Hengge-Aronis, R. (1991). Identification of a central regulator of stationary-phase gene expression in E. coli. Mol Microbiol *5*, 49-59.

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

Lee, C., Lee, S. M., Mukhopadhyay, P., Kim, S. J., Lee, S. C., Ahn, W. S., Yu, M. H., Storz, G., and Ryu, S. E. (2004). Redox regulation of OxyR requires specific disulfide bond formation involving a rapid kinetic reaction path. Nat Struct Mol Biol *11*, 1179-1185.

Leslie, A. G. (1999). Integration of macromolecular diffraction data. Acta Crystallogr D Biol Crystallogr *55 (Pt 10)*, 1696-1702.

Lesniak, J., Barton, W. A., and Nikolov, D. B. (2002). Structural and functional characterization of the Pseudomonas hydroperoxide resistance protein Ohr. Embo J *21*, 6649-6659.

Libby, S. J., Goebel, W., Ludwig, A., Buchmeier, N., Bowe, F., Fang, F. C., Guiney, D. G., Songer, J. G., and Heffron, F. (1994). A cytolysin encoded by Salmonella is required for survival within macrophages. Proc Natl Acad Sci U S A *91*, 489-493.

Lim, D., Poole, K., and Strynadka, N. C. (2002). Crystal structure of the MexR repressor of the mexRAB-oprM multidrug efflux operon of Pseudomonas aeruginosa. J Biol Chem *277*, 29253-29259.

Liu, Y., Manna, A., Li, R., Martin, W. E., Murphy, R. C., Cheung, A. L., and Zhang, G. (2001). Crystal structure of the SarR protein from Staphylococcus aureus. Proc Natl Acad Sci U S A *98*, 6877-6882.

Lonetto, M., Gribskov, M., and Gross, C. A. (1992). The simga 70 family: sequence and evolutionary relationships. J bacteriol *174*, 3843-3849.

Ludwig, A., Bauer, S., Benz, R., Bergmann, B., and Goebel, W. (1999). Analysis of the SlyA-controlled expression, subcellular localization and pore-forming activity of a 34 kDa haemolysin (ClyA) from Escherichia coli K-12. Mol Microbiol *31*, 557-567.

Ludwig, A., Tengel, C., Bauer, S., Bubert, A., Benz, R., Mollenkopf, H. J., and Goebel, W. (1995). SlyA, a regulatory protein from Salmonella typhimurium, induces a haemolytic and pore-forming protein in Escherichia coli. Mol Gen Genet *249*, 474-486.

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.

Mansoor, S. E., and Farrens, D. L. (2004). High-throughput protein structural analysis using site-directed fluorescence labeling and the bimane derivative (2-pyridyl)dithiobimane. Biochemistry *43*, 9426-9438.

Martin, R. G., Jair, K. W., Wolf, R. E., Jr., and Rosner, J. L. (1996). Autoactivation of the marRAB multiple antibiotic resistance operon by the MarA transcriptional activator in Escherichia coli. J Bacteriol *178*, 2216-2223.

Martin, R. G., and Rosner, J. L. (1995). Binding of purified multiple antibiotic-resistance repressor protein (MarR) to mar operator sequences. Proc Natl Acad Sci U S A *92*, 5456-5460.

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

McKay, D. B., and Steitz, T. A. (1981). Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA. Nature *290*, 744-749.

McPherson, A. (1999). Cystallization of biological macromolecues (Cold Spring Harbor, NY, Cold Spring Harbor Press).

Miller, P. F., and Sulavik, M. C. (1996). Overlaps and parallels in the regulation of intrinsic multiple-antibiotic resistance in Escherichia coli. Mol Microbiol *21*, 441-448.

Mongkolsuk, S., Praituan, W., Loprasert, S., Fuangthong, M., and Chamnongpol, S. (1998). Identification and characterization of a new organic hydroperoxide resistance (ohr) gene with a novel pattern of oxidative stress regulation from Xanthomonas campestris pv. phaseoli. J Bacteriol *180*, 2636-2643.

Nakano, S., Kuster-Schock, E., Grossman, A. D., and Zuber, P. (2003). Spx-dependent global transcriptional control is induced by thiol-specific oxidative stress in Bacillus subtilis. Proc Natl Acad Sci U S A *100*, 13603-13608.

Navarre, W. W., Halsey, T. A., Walthers, D., Frye, J., McClelland, M., Potter, J. L., Kenney, L. J., Gunn, J. S., Fang, F. C., and Libby, S. J. (2005). Co-regulation of Salmonella enterica genes required for virulence and resistance to antimicrobial peptides by SlyA and PhoP/PhoQ. Mol Microbiol *56*, 492-508.

Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281-296.

Nikaido, H. (1998). Multiple antibiotic resistance and efflux. Curr Opin Microbiol *1*, 516-523.

Ochsner, U. A., Hassett, D. J., and Vasil, M. L. (2001). Genetic and physiological characterization of ohr, encoding a protein involved in organic hydroperoxide resistance in Pseudomonas aeruginosa. J Bacteriol *183*, 773-778.

Ogata, C. M. (1998). MAD phasing grows up. Nat Struct Biol *5 Suppl*, 638-640.

Oscarsson, J., Mizunoe, Y., Li, L., Lai, X. H., Wieslander, A., and Uhlin, B. E. (1999). Molecular analysis of the cytolytic protein ClyA (SheA) from Escherichia coli. Mol Microbiol *32*, 1226-1238.

Oscarsson, J., Mizunoe, Y., Uhlin, B. E., and Haydon, D. J. (1996). Induction of haemolytic activity in Escherichia coli by the slyA gene product. Mol Microbiol *20*, 191-199.

Pabo, C. O., and Lewis, M. (1982). The operator-binding domain of lambda repressor: structure and DNA recognition. Nature *298*, 443-447.

Paget, M. S., and Buttner, M. J. (2003). Thiol-based regulatory switches. Annu Rev Genet *37*, 91-121.

Petersohn, A., Engelmann, S., Setlow, P., and Hecker, M. (1999). The katX gene from bacillus subtilis is under dual control of sigma B and simga F. Mol Gen Genet *262*, 173-179.

Poole, K., Tetro, K., Zhao, Q., Neshat, S., Heinrichs, D. E., and Bianco, N. (1996). Expression of the multidrug resistance operon mexA-mexB-oprM in Pseudomonas aeruginosa: mexR encodes a regulator of operon expression. Antimicrob Agents Chemother *40*, 2021-2028.

Powell, D. R. (2005). Crystallography notes. http://xraywebmsgkuedu/notes/crystallographyhtml#bragg.

Price, C. W. (2000). Protective function and regulation of general stress response in bacillus subtilis and related gram-positive bacteria. In Bacterial stress response (Washington, D. C., ASM press), pp. 179-197.

Price, C. W., Murphy, C. K., and Youngman, P. (2001). Genemone wide analysis of the general stress response in Bacillus subtilis. Mol Microbiol *41*.

Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature *362*, 219-223.

Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R., and Sklenar, H. (1989). Conformational and helicoidal analysis of 30 PS of molecular dynamics on the d(CGCGAATTCGCG) double helix: "curves", dials and windows. J Biomol Struct Dyn *6*, 669-699.

Rowlett, R. (2005). Protein X-ray crystallography methods.

Saito, K., Akama, H., Yoshihara, E., and Nakae, T. (2003). Mutations affecting DNA-binding activity of the MexR repressor of mexR-mexA-mexB-oprM operon expression. J Bacteriol *185*, 6195-6198.

Scharf, C., Riethdorf, H., Engelmann, U., and Hecker, M. (1998). Thioredosin is an essential protein induced by multiple stresss in bacillus subtilis. J bacteriol *180*, 1869.

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

Schumacher, M. (2004). Protein crystallography lecture note (MAD phasing).

Schumacher, M. A., Allen, G. S., Diel, M., Seidel, G., Hillen, W., and Brennan, R. G. (2004). Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P. Cell *118*, 731-741.

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

Shea, R. J., and Mulks, M. H. (2002). ohr, Encoding an organic hydroperoxide reductase, is an in vivo-induced gene in Actinobacillus pleuropneumoniae. Infect Immun *70*, 794-802.

Snyder, L., and Champness, W. (1997). Molecular Genetics of Bacteria (Washington, D. C., ASM Press).

Spory, A., Bosserhoff, A., von Rhein, C., Goebel, W., and Ludwig, A. (2002). Differential regulation of multiple proteins of Escherichia coli and Salmonella enterica serovar Typhimurium by the transcriptional regulator SlyA. J Bacteriol *184*, 3549-3559.

Srikumar, R., Paul, C. J., and Poole, K. (2000). Influence of mutations in the mexR repressor gene on expression of the MexA-MexB-oprM multidrug efflux system of Pseudomonas aeruginosa. J Bacteriol *182*, 1410-1414.

Stapleton, M. R., Norte, V. A., Read, R. C., and Green, J. (2002). Interaction of the Salmonella typhimurium transcription and virulence factor SlyA with target DNA and identification of members of the SlyA regulon. J Biol Chem *277*, 17630-17637.

Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F., and Matthews, B. W. (1982). Structural similarity in the DNA-binding domains of catabolite gene activator and cro repressor proteins. Proc Natl Acad Sci U S A *79*, 3097-3100.

Storz, G. (2000). Bacterial stress responses, ASM press).

Storz, G., and Hengge-Aronis, R., eds. (2000). Bacterial stress responses (Wanshington, D. C., ASM press).

Storz, G., and Imlay, J. A. (1999). Oxidative stress. Curr Opin Microbiol *2*, 188-194.

Stout, G., and Jensen, L. (1989). X-ray structure determination, A practical guide, Second edition edn, A Wiley-interscience publication).

Sukchawalit, R., Loprasert, S., Atichartpongkul, S., and Mongkolsuk, S. (2001). Complex regulation of the organic hydroperoxide resistance gene (ohr) from Xanthomonas involves OhrR, a novel organic peroxide-inducible negative regulator, and posttranscriptional modifications. J Bacteriol *183*, 4405-4412.

Sulavik, M. C., Dazer, M., and Miller, P. F. (1997). The Salmonella typhimurium mar locus: molecular and genetic analyses and assessment of its role in virulence. J Bacteriol *179*, 1857-1866.

Sulavik, M. C., Gambino, L. F., and Miller, P. F. (1995). The MarR repressor of the multiple antibiotic resistance (mar) operon in Escherichia coli: prototypic member of a family of bacterial regulatory proteins involved in sensing phenolic compounds. Mol Med *1*, 436-446.

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

Tao, Y. (2005). Model Building, Maps and Evaluation of Model Quality (http://biocriceedu/~ytao/Bios482/lecture12ppt).

Terwilliger, T. C. (1994). MAD phasing: treatment of dispersive differences as isomorphous replacement information. Acta Crystallogr D Biol Crystallogr *50*, 17-23.

Terwilliger, T. C. (1999). Reciprocal-space solvent flattening. Acta Crystallogr D Biol Crystallogr *55*, 1863-1871.

Terwilliger, T. C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr *55 (Pt 4)*, 849-861.

Vagin, A., and Teplyakov, A. (1997). Molrep: An automated program for molecular replacement. J Appl Cryst *30*, 1022-1025.

Wada, A. (1976). The alpha-helix as an electric macro-dipole. Adv Biophys, 1-63.

Wallace, A. J., Stillman, T. J., Atkins, A., Jamieson, S. J., Bullough, P. A., Green, J., and Artymiuk, P. J. (2000). E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. Cell *100*, 265-276.

Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J., and Sussman, J. L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. Proc Natl Acad Sci U S A *97*, 623-628.

White, A., Ding, X., vanderSpek, J. C., Murphy, J. R., and Ringe, D. (1998). Structure of the metal-ion-activated diphtheria toxin repressor/tox operator complex. Nature *394*, 502-506.

Wilkinson, S. P., and Grove, A. (2006). Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. Curr Issues Mol Biol *8*, 51-62.

Wintjens, R., and Rooman, M. (1996). Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. J Mol Biol *262*, 294-313.

Wu, R. Y., Zhang, R. G., Zagnitko, O., Dementieva, I., Maltzev, N., Watson, J. D., Laskowski, R., Gornicki, P., and Joachimiak, A. (2003). Crystal structure of Enterococcus faecalis SlyA-like transcriptional factor. J Biol Chem *278*, 20240-20244.

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.

Zheng, M., Aslund, F., and Storz, G. (1998). Activation of the OxyR transcription factor by reversible disulfide bond formation. Science *279*, 1718-1721.

Zheng, N., Fraenkel, E., Pabo, C. O., and Pavletich, N. P. (1999). Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. Genes Dev *13*, 666-674.