STRUCTURAL STUDIES OF THE LYSR-TYPE TRANSCRIPTIONAL REGULATOR BENM DNA BINDING DOMAIN WITH ITS DNA REGULATORY REGIONS

By

Amer Alanazi

(Under the direction of Cory A. Momany)

ABSTRACT

The LysR-type transcriptional regulator (LTTR) BenM is involved in controlling benzoate degradation in the soil bacterium *Acinetobacter baylyi* strain ADP1. The 1.8 Å resolution crystal structure of the unbound BenM DNA binding domain (BenM DBD) confirmed that the BenM DBD forms a compact globular domain composed of three helices with helix-turn-helix motif (α2-α3), a wing, and a long linker-helix. BenM DBD was crystallized with short oligonucleotides from its cognate *benA* and *catB* DNA promoters in two different crystal-packing arrangements. In these nucleic acid complexes, BenM DBD dimers span a large region of bent DNA where the DNA recognition helices (α2) of one dimer bind into two consecutive DNA major grooves in a sequence-dependent manner. The specific DNA major groove interactions that define the LTTR conserved recognition motif (T-N_{11}-A) include van der Waals interactions of two proline residues at the N-terminal end of the recognition helices with the methyl groups of thymine bases within the recognition motif. Also involved in sequence specific

interactions are the side chain of Gln 29 with the imino and amino groups of an adenine base respectively (5'-ATAC-3') and the side chain of Arg 34 with the carbonyl oxygen of guanine (5`-GTAT-3`) in the complementary strand. A wing interacts mainly with the phosphate backbone of the DNA minor groove and assists in the proper positioning of the N-terminal end of the recognition helix.

The BenM DBD/catB site 1 complex demonstrated very tight crystal packing with face to face contacts similar to the natural protein-protein interactions. The arrangement of the two DNA duplexes was unique. Chains E and F run as two columns parallel to the crystallographic *c* cell axis and contact one of the symmetry related protein subunits. Chains G and H create an end-to-end continuous helix with other symmetry-related mates and run perpendicular to the crystallographic *c* cell axis. The linker helices of BenM DBD chains A and C cross one-another, whereas linker helices of B and D run parallel to A and D respectively. These unique crystal packing arrangements might be generated and perhaps stabilized by the association of two malonate molecules at the N-terminal ends of two monomers each from dimer.

Index words: LysR-type transcriptional regulators, BenM, DNA binding domain, protein-DNA interactions.

STRUCTURAL STUDIES OF THE LYSR-TYPE TRANSCRIPTIONAL REGULATOR BENM DNA BINDING DOMAIN WITH ITS DNA REGULATORY REGIONS

By

Amer M Alanazi

B.S., King Saud University, Saudi Arabia, 2003

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

STRUCTURAL STUDIES OF THE LYSR-TYPE TRANSCRIPTIONAL REGULATOR BENM DNA BINDING DOMAIN WITH ITS DNA REGULATORY REGIONS

By

Amer M Alanazi

Major Professor: Cory Momany

Committee: Warren Beach

Shelly Hooks

Robert Phillips

Ellen Neidle

Electronic version approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
Dec 2010

DEDICATION

This work is dedicated to my dear parents for their unconditional support and encouragement, to my lovely intimate life companion Bodor, for her patience and great sacrifice, and lastly but not the least, to my adorable children, Aban and Retal.

ACKNOWLEDGEMENTS

I would like to thank first and foremost my God “ALLAH” for his countless blessings and guidance throughout my entire life. Second, I want to express my gratitude to my major professor, Dr. Cory Momany, who taught me and helped me in many aspects throughout my graduate studies. I am deeply grateful, thankful and appreciative for all of his cooperation, passion, research support and supervision. I am thankful to Dr. Ellen Neidle for her advice and comments that helped me to proceed and overcome obstacles through my research. I also extend acknowledgment to Dr. Shelly hooks, Dr. Robert Phillips, and Dr. Warren Beach for their time, comments and guidance during my committee meetings.

Finally, I would like to thank the Saudi Arabian Cultural Mission for their financial support of my graduate research.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

The central dogma of molecular biology is that genetic material encoded by DNA is transcribed into mRNA and then translated into protein. Gene transcription and translation together result in gene expression. In the prokaryotes, there are various environmental stimuli that can affect the metabolism, transcription and regulation of bacteria in their response for survival. One response towards those external factors is by regulation of gene expression[1, 2]. To regulate gene expression at the transcriptional level, bacteria utilize both repression of transcription and activation of transcription. LysR-type transcriptional regulators (LTTRs) are one class of proteins that can perform both functions. Studies on these proteins are the focus of this dissertation.

LysR-Type Transcriptional Regulators (LTTRs)

Transcriptional regulators can be categorized into many families with new families constantly emerging[1, 5]. One of the most common and divergent families is the LysR-type transcriptional regulator family. LTTRs were first reported by Henikofff in 1988[6]. Members of this family regulate many metabolic functions, involved with virulence factors, CO_2 fixation, nodulation, host-microbe interactions and antibiotic resistance. LTTR proteins regulate many complex genes,

autoregulate their own transcription and also negatively regulate genes where they work as transcriptional repressors. Amazingly, despite regulating many diverse genes, they show sequence conservation of only 20% or more at the full-length level. Most of the highest sequence conservation lies at the N-terminal end of the protein where there is a conserved helix-turn-helix DNA binding motif. LTTRs share some common characteristics: they bind to small molecular ligands, then repress their own transcription to maintain consistent levels, and they bind to their target DNA independently of the ligand binding[7, 8]. The family members usually range from 300-350 amino acids in sequence and assembles into homodimers or homotetramers for DNA binding activity. They have two main domains: The N-terminal domain contains the helix-turn-helix motif, which is well conserved motif among the family members of DNA-binding proteins. Therefore, it is responsible for DNA binding of LTTRs and is referred to as the DNA binding domain (DBD). The other domain is the C-terminal domain which is responsible about the ligand binding, so it is known as the effector binding domain (EBD). This domain has less sequence conservation, but is structurally conserved among the family members.

LTTRs perform transcriptional regulation by controlling access of a multi-subunit DNA-dependent RNA polymerase to the DNA in the region termed a “promoter”. The promoter is the site where RNA polymerase begins RNA synthesis. Within prokaryotic promoters, there are several relatively well-conserved elements that have been identified. The main elements are the -10 region (the Pribnow box), and -35 region, which are located 10 and 35 base pairs

(bp) upstream from the transcript start site. In addition, there can be extended -10 and UP elements. Each one of these elements is recognized by subunits of the RNA polymerase. The core RNA polymerase enzyme has five subunits: $\beta\beta'\alpha_2\omega$. Domain 2 of the RNA polymerase σ subunit recognizes the -10 hexamer regions; while domain 4 recognizes the -35 hexamer region. The extended -10 region is a small motif (3-4 bp) located immediately upstream of the -10 region and is recognized by the domain 3 of the σ subunit. Lastly, the UP elements are recognized by the C-terminal domain of the α subunit (αCTD).

Once the RNA polymerase binds to DNA, a "bubble" is generated by unwinding of the DNA at the transcript start site. One of the most important mechanisms of regulation is the control of the initial binding of the RNA polymerase enzyme to the target promoter sequence[3]. This is a common mode used by LysR-Type Transcriptional regulators (LTTRs).

General features of LTTR binding with DNA

One of the unique characteristics of this family of proteins is that they can bind to their target promoters with or without the presence of effectors or co-inducers. Binding of the co-inducers causes conformational changes that dictate which transcriptional state the regulator is in[9]. The co-inducers of many LysR-regulators are usually metabolic intermediates in the corresponding metabolic pathway that they regulate. The changes in the expression patterns of the enzymes are dependent on the concentrations of the co-inducer. The DNA that encodes LTTRs is frequently organized in a transcriptionally divergent orientation

with respect to the genes they regulate. This allows them to negatively regulate their own expression and positively regulate the expression of their target genes upon co-inducer binding. Interactions of LTTRs with DNA can be complicated because they can form multimeric oligomers and occupy multiple binding sites on the DNA promoter[10]. Moreover, binding of their cognate co-inducers often generates quantitative and qualitative changes of the DNA-protein complex that can increase the number of bound promoters[10]. DNA protection patterns from foot printing studies showed that LTTRs can protect large regions of DNA promoter. There are two regions that those regulators bind: One region is used for repression and is called the “repression binding site” (RBS). The other region is used for activation and is called the “activation binding site” (ABS)[9, 11]. DNase foot printing studies of the LTTR BenM with the *benA* promoter DNA showed a large protected region by BenM. The large protected region consists of three sites. The first site overlaps with the RBS of LTTR binding and is occupied by the tetrameric protein when there is no co-inducer bound. This region contains the conserved LTTR sequence, T-N_{11}-A that is known as the “LTTR-recognition motif”[8, 18]. BenM protein in the tetrameric form in the absence of co-inducer occupies both site 1, which contains the specific and conserved LTTR motif for BenM protein: ATAC-N_7-GTAT with two dyad symmetry, and site 3, which differs from site 1 in one base, ATAC-N_7-GTAT. When the co-inducers, benzoate and/or *cis*,*cis*-muconate, bind to BenM the site 3 region is no longer protected and the protein shifts to site 2. The center of site 2 is separated by 21 bp from the center of site 1. This shifting is due to conformational changes of the tetrameric BenM

protein and consequently allows for the RNA polymerase to bind in that region (site 3) to activate the transcription of the *benA* promoter. Site 1 is protected in all cases, whether there is co-inducer bound or not. Association of BenM with Site 1 causes transcriptional repression of BenM since site 1 overlaps with the RBS[11]. Similar findings have been found with CatR and CclR when binding to their promoter recognition sites[12]. CysB protein from *E. coli* exhibited a similar DNA binding pattern as BenM protein from *A. baylyi* ADP1. When the inducing ligand N-acetyl cysteine (NAC) of CysB is absent, CysB tetramer binds to its positively regulated genes and occupies multiple binding sites. Therefore, transcriptional activation of CysB is dependent on binding of NAC, which causes conformational changes and DNA bending[9]. Another level of transcriptional regulation has been shown for the redox sensitive regulator, OxyR. Multiple active forms of OxyR when bound to DNA produced unique alterations in the DNA structure where each one of those active states of the regulator had different affinity and consequently different mechanisms of transcriptional regulation. These different active states of OxyR are due to modifications of the cysteine residues in response to redox changes that produce transitions of the oligomeric state from dimer to tetramer[13]. Another common mechanism of LTTR binding sites that has been seen with many LTTRs is the DNA bending upstream from the -35 box of the activated promoter region. This bending might be due to the arrangement of the four DNA binding domains of the tetrameric form of LTTRs[14, 15, 16, 17].

DNA recognition by the helix-turn-helix motif and its structural variations

The study of DNA binding proteins is an essential area of research as they play such an important role in molecular biology. They are responsible for replicating the genome, for transcribing genes in active loci, for repairing damaged DNA, for packaging and rearrangement of chromosomes. DNA-binding proteins can be classified based on the structures of the DNA binding domain regions[19, 20]. Among many DNA binding motifs, the helix-turn-helix (HTH) motif is the most predominant and best-characterized motif used by prokaryotes to bind DNA. The classic arrangement of this motif is two helices arranged almost perpendicular to each other joined by a tight turn, usually three amino acids in length. The HTH motif alone is insufficient to stabilize the fold, so a third helix is joined via a turn to make a rigid globular domain. The secondary structural arrangement of the motif is as follows: helix 1, turn, helix 2, another turn, helix 3. Helix 2 and 3 define the HTH motif and helix 3 is known as the DNA recognition helix. The motif can be found as part of a bundle of 3 to 6 helices, which together provides a very stable domain. The recognition helix is important for specificity and is broadly found to make extensive and specific DNA base pair contacts to the major groove[21]. These specific interactions are formed by a network of hydrogen bonds and other contacts between exposed bases within the DNA major groove to the side chains of specific amino acids in the recognition helix of the DNA binding protein[22]. Additional supporting contacts with the DNA backbone can come from the linkers and first helix of the motif. There have been many variations found in the classical HTH motif where the turn between the two

helices of the motif is longer and adopts different conformations. Other variations have been determined in the topology of the helices. The first HTH domains of protein-DNA complexes that were characterized were completely α helical. Later structures showed that β strands could interrupt, precede or follow the α helices involved in DNA binding. DNA binding domains made solely of β-strands have now been added to the list. In some, the β strands have been found to be packed against the motif helices by forming an extended antiparallel β-sheet. Since these structural features look like a “wing”, the variants have been called “winged” helix proteins[23, 24]. As more HTH domain-DNA structures have become available, many variations have been found in the motif in different protein-DNA complexes[25] (Figure 1.1). Thus there is diversity in the HTH DNA-binding domains with regard to the function, topology and conformation of the turn between the two helices[24]. Recently many structures of eukaryotic homeodomains and transcriptional factors have been determined that demonstrate the HTH motif.

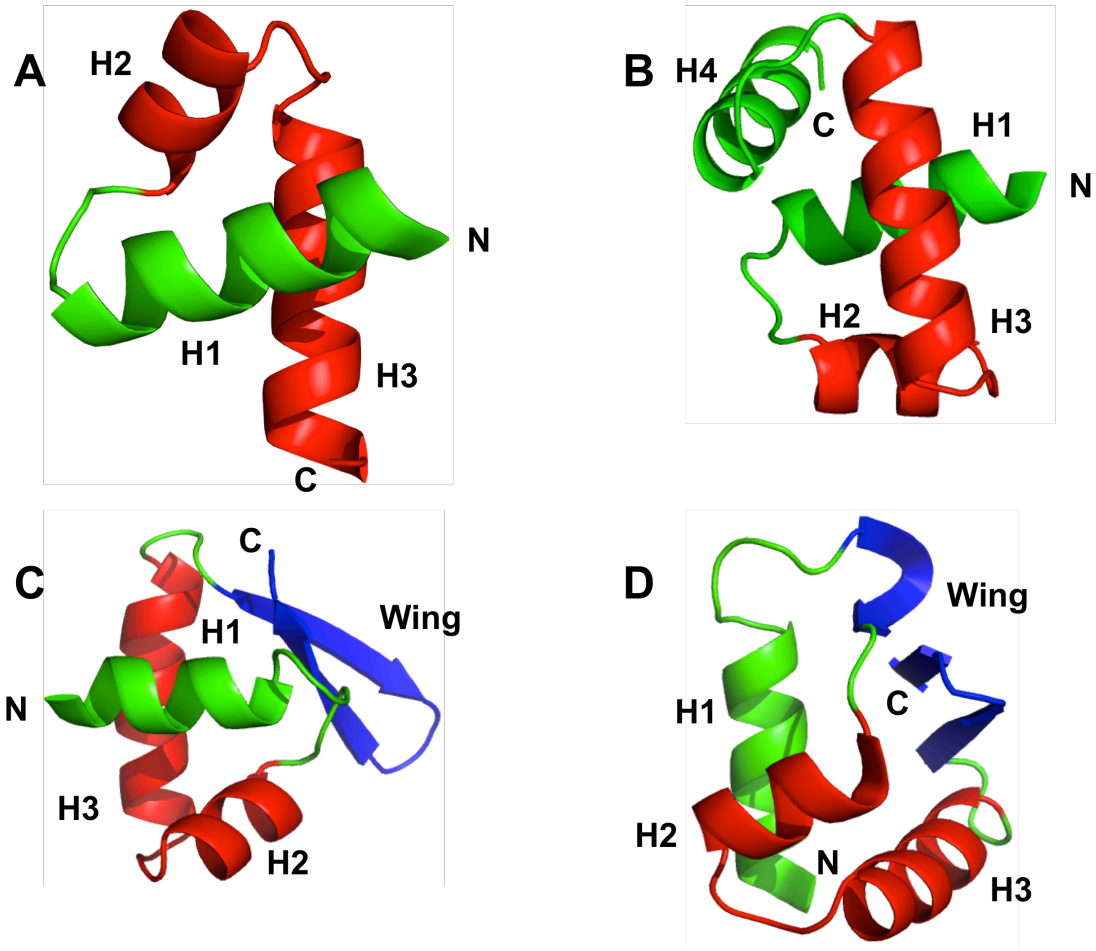


Figure 1.1. Ribbon representation of the HTH motif and its variants. The helices are labeled H and shown as green, the HTH motif is shown in red (H2, H3), and the β strands are in blue. (A) Simple tri-helical HTH motif (PDB: 1K78), (B) tetra-helical bundle (PDB: 1A04), (C) simple 2-stranded winged HTH (PDB: 1SMT), (D) 4-stranded winged HTH (PDB: 1CGP).

With the exception of AraC family members Rob and MarA from *E. coli* that use monomers of the HTH motif to bind to their cognate DNA binding sites, almost all prokaryotic transcriptional regulators use the HTH motif to bind as homodimers to their palindromic DNA binding sites where each motif binds to one half-site of symmetry-related DNA[21]. On the other hand, eukaryotic transcriptional regulators bind with different oligomeric states such as monomers or heterodimers and to non-symmetrical DNA sites. Therefore, these different conformations of the motif allow for a wide recognition range of DNA sequences[21]. Rob and MarA are activating genes involved in antibiotic and heavy metal resistance. The striking structural difference between the monomers of Rob and MarA is that, the Rob monomer has two HTH motifs, and yet only one of them is inserted into the major groove where it makes multiple base specific contacts. The other HTH contacts the minor groove backbone only. In contrast, the two HTH motifs of MarA monomer bind into two consecutive major grooves[21, 26, 27] (Figure 1.2).

The first winged helix protein structure to be discovered came from the structure of the DNA binding domain of the hepatocyte nuclear factor 3γ (HNF-3γ) with its promoter target site[23, 28]. Since then, more winged HTH motif structures have become available. The motif is a derivative or subfamily of the HTH super family and forms as a compact α/β structure where the N-terminal part is mainly α helical, the C-terminal part contains 2 to 4 β-strands flanked by wings. The turn between the two helices of the HTH motif can vary in length

allowing for different angle positions over the ~120° angle seen with the canonical HTH proteins[29].

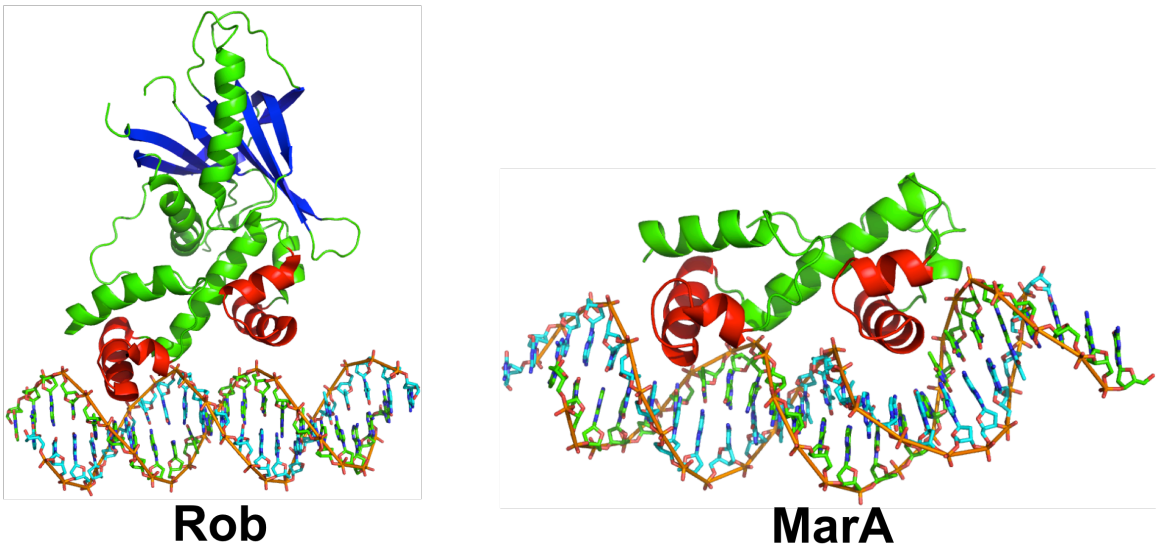


Figure 1.2. Ribbon representation of AraC family members Rob and MarA monomers are shown. The HTH motif is colored red, α-helices are colored green and the β-strands are colored blue. The DNA duplexes are shown as sticks.

The DNA recognition of those winged helix proteins is similar to that of the classical HTH proteins where the recognition helix is still inserted in the major groove. However, both β-strands and the wings make polar contacts directly with the minor groove or through water molecules found at the interface between the protein and the DNA. Another role for the wings has been seen in the structure of the E2F family transcriptional factor with its DNA complex. The wings of the E2F factor are not at the center of DNA binding; instead they are located far away

from the protein-DNA interface where they mediate protein-protein interactions as well as dimerization between the helices owing to their flexibility. Therefore, winged helix proteins can function to stabilize the fold, in addition to making sequence specific contacts to DNA as in the HTH proteins[29, 30].

The structure of the bacteriophage Mu repressor (MuR) DNA binding domain with its DNA complex showed that the wing is immobilized in the minor groove through intermolecular hydrogen bonding contacts and assists in the proper positioning of the motif on DNA (Figure 1.3) [31]. BmrR transcriptional regulator from *B. subtilis* was the first MerR family member structure solved as a DNA complex, and its structure explained the ligand-induced activation of this family. BmrR is involved in activating multi-drug transporters upon binding to lipophilic cations, which then causes multi-drug resistance. The structure showed that BmrR binds as a homodimer to its promoter and contains three domains: a winged HTH domain, a long helical linker formed by antiparallel coiled coil, and finaly a drug-binding domain that dimerizes with the DNA binding domain. The helices of the HTH motif make contacts with two consecutive major grooves. The wing positions the domain on the DNA makes contacts with the DNA phosphate backbone and sugar moieties[21, 32] (Figure 1.3).

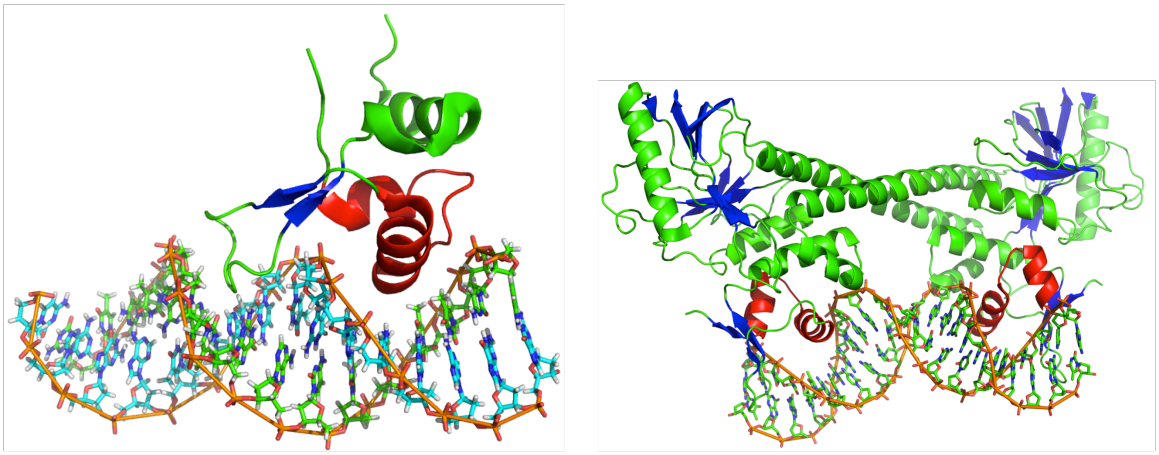


Figure 1.3. Ribbon representation of the bacteriophage Mu repressor (MuR) and the MerR family member BmrR are shown. The coloring scheme used is as in the previous figures.

The genes involved in fatty acid metabolism of *E. coli* are activated and repressed by the GntR family transcriptional regulator FadR upon binding to long chain Acyl-CoA molecules. The DNA bound-FadR structure showed that the protein is a homodimer with two domains, a winged helix DNA binding domain and a C-terminal Acyl CoA-binding site domain. Surprisingly, the two recognition helices of the dimer do not occupy two consecutive major grooves; instead they were deeply inserted into the same major groove with two central base pairs separating them. Just like in the Mu repressor, the wings were inserted into the minor groove and making contacts with the DNA backbone[21, 33] (Figure 1.4). Tetracycline resistance in Gram-negative bacteria is regulated by TetR repressor when it binds to its *tetO* operator half sites. The DNA-complex structure of TetR with its ligand Mg^{2+} provides new features of DNA recognition for the winged helix

family proteins. The TetR-DNA interface had no water molecules and no minor groove contacts, unlike the structures of other family members. Moreover, there were no major groove contacts to the central base pairs as well, which was necessary to allow for proper spacing between the *tetO* operator half sites. A more distinguishable structural feature is the widening of the two consecutive major grooves (~14.5 Å versus 11.7 Å) for canonical B-DNA. Such widening has been associated with unwinding of the DNA[21, 34] (Figure 1.4).

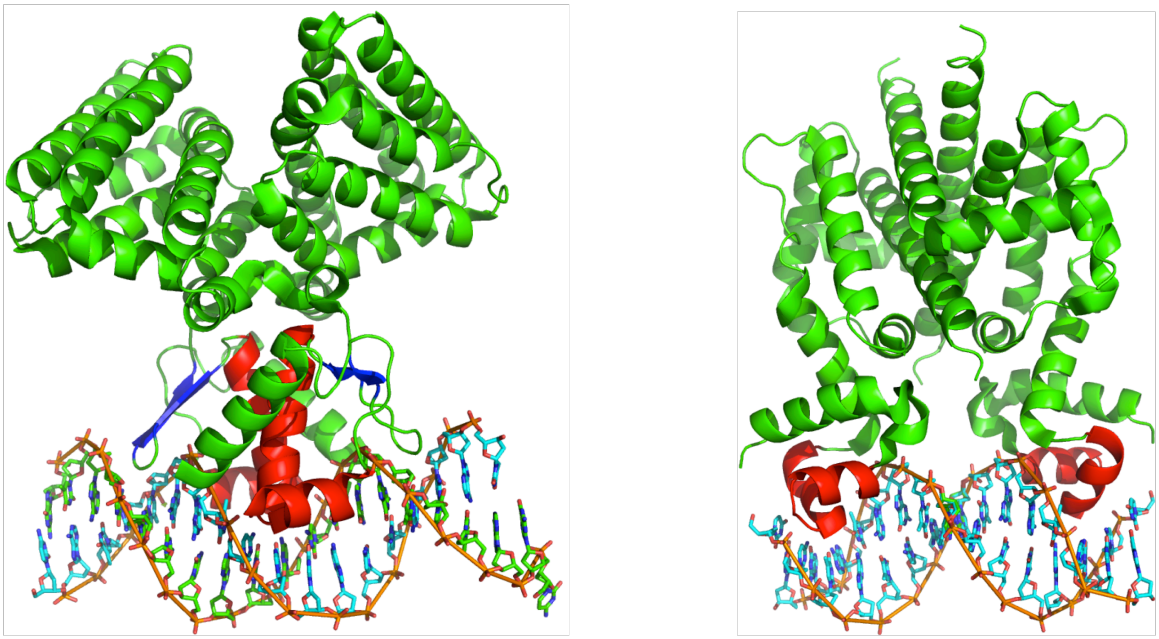


Figure 1.4. Ribbon representation of the GntR family member FadR and the tetracycline resistance repressor TetR are shown. The coloring scheme is as in the previous figures.

Based on the protein-DNA complexes presented here from different families of transcriptional regulators, it is clear that the classic and winged HTH motifs can be varied structurally to accommodate many different DNA recognition mechanisms necessary for activating and repressing genes involved in metabolic and cellular functions.

Structural studies of LTTRs

LTTRs can be organized into two domains, a DNA binding domain (DBD) at the N-terminus and a effector binding domain (EBD) at the C-terminus. The domains are connected by a long linker helix[35, 36] (Figure 1.5). These structural features have been observed in multiple EBDs and full-length LTTR structures. The DBD uses either a winged HTH or just a classic tri-helical motif. The EBD has a periplasmic–binding protein fold, which consists of the conserved Rossmann-like fold[37]. The EBD has two similar α/β domains, I and II, which are connected by two cross-over β-strands.

Owing to the great diversity of LTTRs, structural characterization and analysis of the proteins have proceeded slowly. Successful structural studies of LTTRs started with truncated versions of the protein where the DBD along with the linker helix was removed to improve the solubility of the purified protein. Inclusion of the DBD on the full-length proteins has been shown to be associated with insolubility problems that can occur during purification or crystallization of the full-length proteins.[39] The DBD is speculated to undergo domain swapping, where DBDs interact with dimers from neighboring molecules rather than the

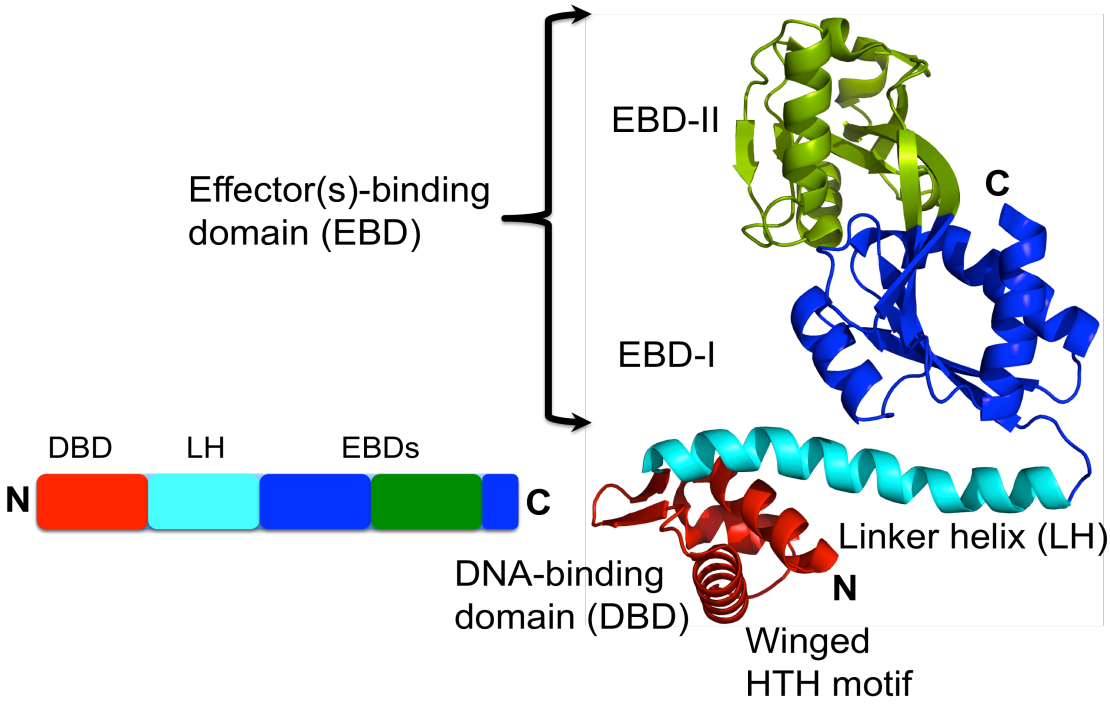


Figure 1.5. Structural organization of the LTTR domains as shown in the monomeric full-length structure of CbnR (PDB: 1IZ1).

functional oligomeric unit.[38] The first LTTR structure solved was the CysB cofactor-binding domain (EBD) from *Klebsiella aerogenes* because no suitable full-length crystals could be obtained.[39] CysB controls genes responsible for the biosynthesis of cysteine upon binding by the inducer N-acetylserine. Two sets of mutations were conducted to map out the functions of these protein's domains. Mutation of Ser 34 to arginine, which is highly conserved among LTTRs at the structurally equivalent positions, caused loss of DNA binding while mutation at the dimer interface between the two monomers caused loss of cofactor binding. The CysB cofactor-binding domain contains two α/β domains linked by two antiparallel β-strands[39] (Figure 1.6). Soon following CysB was another truncated

LTTR structure (EBD), the OxyR C-terminal domain from *E. coli* in both reduced and oxidized states. OxyR is responsible for activating genes that defend the organism against reactive oxygen species that cause oxidative stress. The protein undergoes a conformational "switch" by H_2O_2 oxidation, which induces the formation of an intermolecular disulfide bond between the redox-active cysteines Cys 199 and Cys 208. The C-terminal domain of OxyR was similar in structure to CysB, having two α/β domains connected by two inter-domain with β-strands. The two reactive cysteines are positioned far from one another in the reduced state, with Cys 199 positioned in a small hydrophobic pocket at the interface between the two domains and Cys 208 at the lower part of domain II (Figure 1.6).[40] DntR is another LTTR that controls the transcription of enzymes responsible for the oxidative degradation of nitroaromatic compounds such as 2,4-dinitrotoulene in *Burkholderia* sp. strain DNT. The DntR inducer binding domain (IBD) crystallized as a homodimer in the asymmetric unit in a head to tail arrangement of the two monomers. Although the protein crystallized as a full length protein, because of very poor electron density features in the DBD region, the DBD could not be modeled to a level that could be refined. Just like CysB and OxyR, the EBD has two sub-domains and contains a Rossmann-like fold with a hinge region linking those two sub-domains (Figure 1.6).[41]

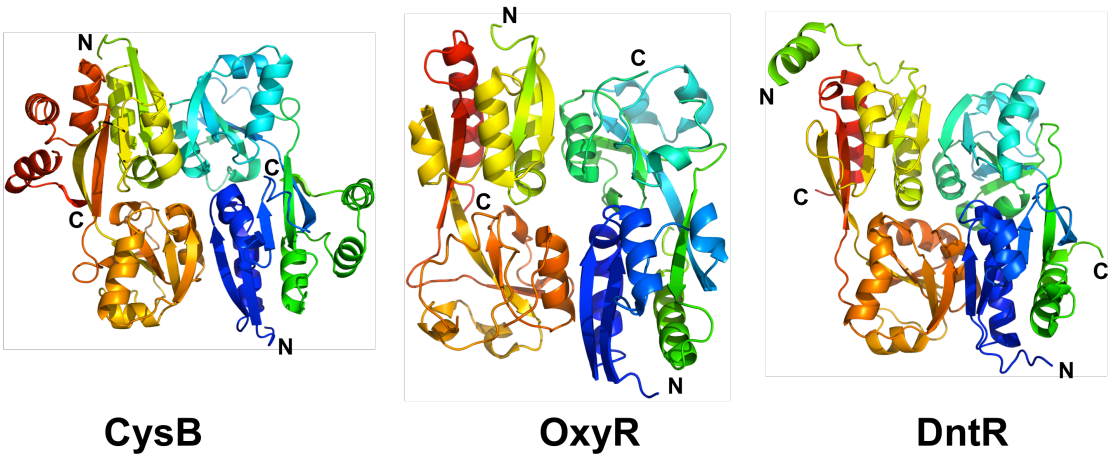


Figure 1.6. Ribbon representations of CysB cofactor-binding domain dimer (PDB: 1AL3), OxyR reduced C-terminus domain dimer (PDB: 1I69), DntR inducer-binding domain dimer (PDB: 1UTB) are shown.

A structure of the EBD of AmpR from *Citrobacter freundii*, which controls the expression of AmpC, has been solved recently.[42] AmpC is involved in the hyperproduction of β-lactamase enzyme, which is responsible for the resistance mechanism seen with the administration of β-lactam antibiotics such as penicillins and cephalosporins in human pathogens such as *Pseudomonas aeruginosa* and Enterobacteriaceae. AmpR activates AmpC upon the cytosolic accumulation of peptidoglycan metabolites that are produced during β-lactam resistance. The overall structure is just like the previously solved LTTR EBDs, it has two monomers, and each has two α/β Rossmann-like fold subdomains (I, II) that form the dimeric active form[42].

There are currently an increasing number of full-length LTTR structures available in the protein data bank from different bacteria and with different

regulatory functions. The following are all the available full-length LTTR structures: CbnR from *Ralstonia eutropha* NH9 being the first one to be solved[35], TsaR from *Comamonas testosteroni* T-2[43], CrgA from *Neisseria meningitides*[44], ArgP from *Mycobacterium tuberculosis*[45], two other uncharacterized structures (PDB: 2ESN and 3FZV both from *Pseudomonas aeruginosa*), and recently wild type and two variants of BenM from *Acinetobacter baylyi* ADP1[46]. CbnR activates genes responsible for the degradation of chlorocatechol. The structure is a homotetramer arranged as a dimer of dimers, which is proposed to be the biological active form. The subunits of the tetramer assume two conformations: a compact and an extended conformation. Contacts between LHs and EBD of the two dimers at the tetramerization interface create a closed tetramer (Figure 1.7). Each subunit contains a N-terminal HTH DNA binding domain and a C-terminal effector binding domain (Figure 1.5). The structure showed for the first time where the DBDs are positioned relatively to the EBDs. The four DBDs were located at the bottom end of each tetramer making a V-shape. The arrangement of all four DBDs of the tetramer on one surface places them such that they can interact with a long piece of DNA promoter. The DBD contains three α-helices, two β-strands and shows some flexibility to a degree that the wing region was not visible in the electron density of one of the subunits. The following residues have been well conserved among different LTTRs in the HTH motif region: Ala22, Gln29, Pro30, Arg34, Gln35, Leu39, Glu40, Leu43, Gly44, Leu47 and Arg50. Other residues Leu39, Leu43, Leu47 are clustered and make a hydrophobic core of the DNA-binding domain. Polar residues located on the surface of the motif

were proposed to be important for DNA binding[35]. Structures of TsaR ligand-free and complexed with *para*-toluenesulfonate (TSA) showed a new open conformation of the tetrameric form due to the lack of contacts between the EBDs of the two dimers. TsaR regulates metabolism of the widely used pollutant TSA in chemical industry. The asymmetric unit of the structure contains a dimer that adopts both compact and extended conformations. Therefore, the structure is different from CbnR in the tetramer form[43] (Figure 1.7). CrgA is involved in host-pathogen interactions in *Neisseria meningitides*. The CrgA structure showed a totally different oligomeric assembly while still maintaining the same fold of both domains common for all LTTRs. The assembly is very unique and shaped like a square or connected to each other to form a closed ring (Figure 1.7.). The ring assembly is formed by contacts via LHs and EBDs of each monomer with two other adjacent monomers. CrgA monomers have only one compact conformation. The extended conformation is not seen unlike the previous structures[44]. ArgP activates the genes important for arginine transport in *Mycobacterium tuberculosis*. The structure was solved as a homodimer like CbnR and TsaR, but adopts an open tetramer conformation just like TsaR because there are no contacts between the two EBDs of the two dimers. However, the two dimers adopt compact and extended conformation like the other previous structures[45]. BenM activates genes responsible about the aromatic degradation of benzoate into catechol upon binding of its inducers, benzoate and *cis,cis*-mconate in *Acinetobacter baylyi* ADP1[46]. All the three full-length solved structures: BenM wild-type, BenM R156H variant, and BenM

E226K variant showed a new oligomerization scheme different than the schemes seen with previous LTTR full-length structures. BenM adopts a similar oligomeric form of CbnR where the four subunits contact via EBDs to form a closed

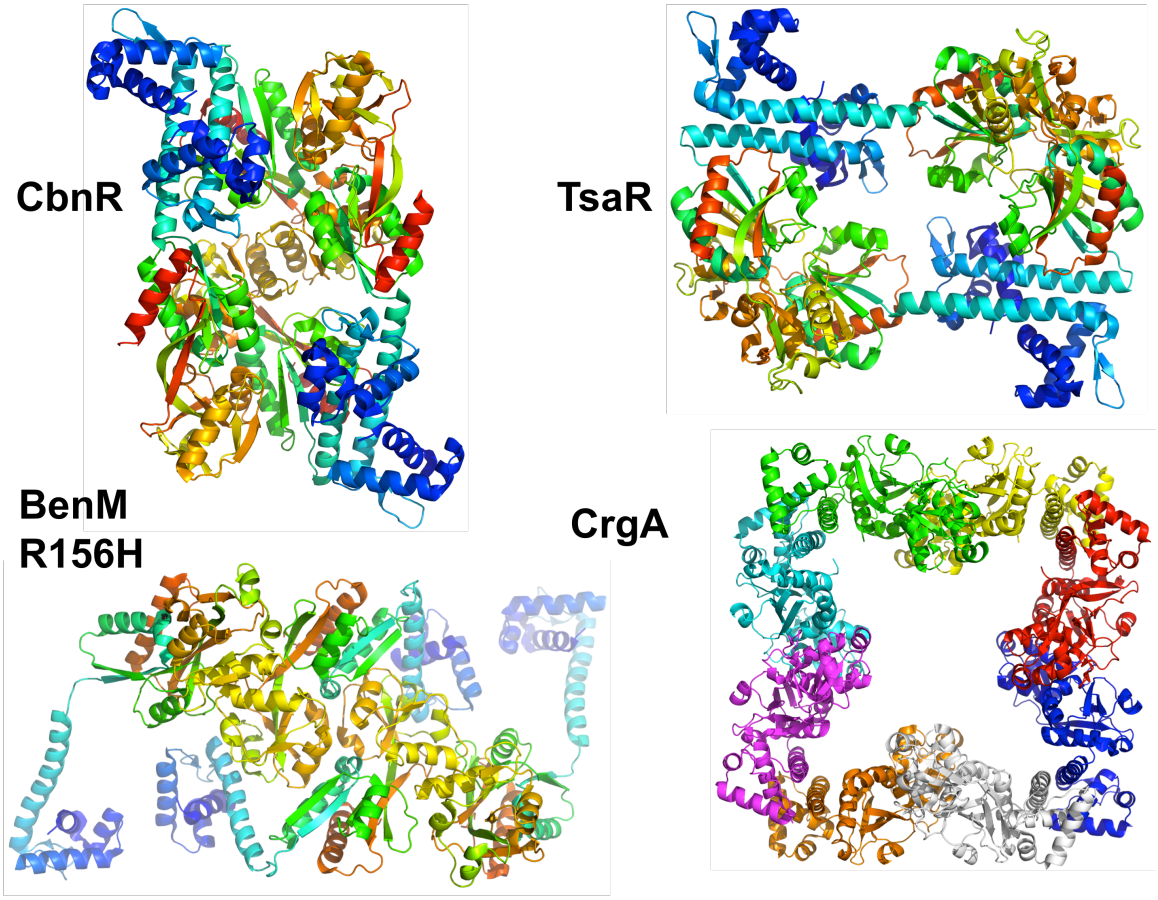


Figure 1.7. Ribbon representation of LTTR full-length tetramers. (PDB codes: CbnR; 1IZ1, TsaR; 3FXQ, BenM R156H; 3K1M, CrgA; 3HHG). All subunits are colored from blue to red, N-terminus to C-terminus except CrgA, each of the eight subunits differently colored.

conformation. However, due to domain swapping of DBDs and varied relationships between LH with DBD/EBD, the closed oligomer conformation cannot be formed[46] (Figure 1.7).

BenM transcriptional regulation in *Acinetobacter baylyi* ADP1

One branch of aromatic compound degradation, the β-ketoadipate pathway, is under the regulation of two LTTR paralogs, BenM and CatM in the soil bacterium *Acinetobacter baylyi* ADP1 (Figure 1.8)[47]. BenM and CatM are highly similar in amino acids sequence, especially in the first 60 amino acids of the N-terminal DNA-binding domain[48]. They both activate the transcription of their target genes upon binding of the same metabolite, *cis, cis*-muconate, which is an intermediate in the degradation of benzoate. However, BenM can also respond to benzoate as a sole effector[47, 49]. Tricarboxylic acid (TCA) intermediates are produced by the action of enzymes encoded by the chromosomal *ben* and *cat* genes[50].

BenM activates the transcription from benA promoter whereas CatM activates the transcription from catB promoter47, 49, 51. Both proteins work equally well at the *catA* promoter. Consistent with other LTTRs, they both transcribe their own genes divergently from the genes they regulate[7]. Interestingly, synergistic transcriptional activation at the *benA* promoter by BenM was observed when both ligands were bound in BenM[11]. However, CatM does not demonstrate this effect. The mechanisms by which these ligands can cause

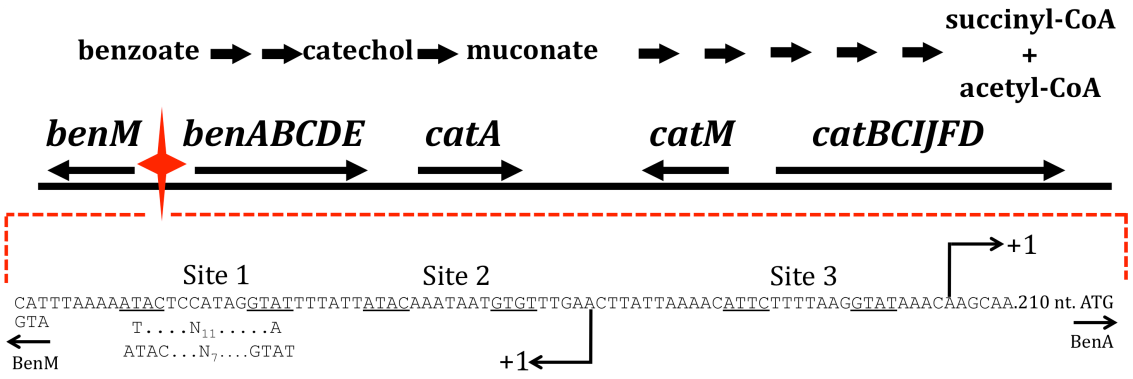


Figure. 1.8. Overview of the β-ketoadipate pathway reactions, gene organization, and BenM binding sites associated with the *benA* promoter. The top line identifies some of the small molecule intermediates involved in benzoate conversion to TCA cycle intermediates. The second line shows the gene organization. The half-sites associated with the three binding sites for both BenM and CatM are underlined in the DNA sequence that is shown for the *benA* promoter with the transcription start sites denoted for BenM and BenA.

this synergistic transcriptional activation have been explained by structural and mutational studies of the BenM effector-binding domain, wild type and variants, with and without effectors. These structures revealed a second binding site (the secondary binding site) that can accommodate benzoate that is distant from the muconate binding site that lies at the center of the EBD. Therefore, the surprising synergistic transcriptional activation effect was due to the simultaneous binding of both benzoate and muconate[48, 52].

All of these structural studies have provided insights on how the diverse family of proteins can interact and how subunits associate to make each

oligomer. Studies also predicted where the ligand binding pockets are and have demonstrated how the conformation of the EBDs change upon ligand binding. However, no structural studies have shown how LTTRs bind to DNA. Therefore, the DNA recognition mechanism is still unknown. The need to unveil the molecular mechanism of LTTR binding to the promoter is imperative. In chapter 2 of this dissertation, structural studies of unbound BenM DBD and with two other regulatory DNA regions are presented and explained in detail. We have also compared the unbound and bound form of BenM and structurally aligned the bound form to other available LTTR DBD structures. The chapter also describes the crystallization screening strategy that we have developed for protein-DNA complexes. In chapter 3, we focus on one of the DNA complexes in terms of crystal packing and demonstrated the uniqueness and oddness of that complex. The chapter also talks about the use of crystallization precipitants or reagents to engineer a DNA-based crystal lattice. The last chapter includes the key findings, conclusions and future directions of the overall research.

References

1. van Hijum, S., M. Medema, et al. (2009). Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiology and Molecular Biology Reviews* **73**: 481-509.
2. Geszvain, K. and R. Landick (2005). The structure of bacterial RNA polymerase. *The bacterial chromosome*: **283**:296.
3. Browning, D. and S. Busby (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2**: 57-65.
4. Ross, W., K. Gosink, et al. (1993). A third recognition element in bacterial promoters: DNA binding by the a subunit of RNA polymerase. *Science* **262**: 1407-1413.
5. Pareja, E., P. Pareja-Tobes, et al. (2006). ExtraTrain: a database of Extragenic regions and Transcriptional information in prokaryotic organisms. *BMC microbiology* **6**: 29.
6. Henikoff, S., G. Haughn, et al. (1988). A large family of bacterial activator proteins. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 6602 .
7. Schell, M. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annual Reviews in Microbiology* **47**: 597-626.
8. Maddocks, S. and P. Oyston (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154**: 3609.

9. Lochowska, A., R. Iwanicka Nowicka, et al. (2004). Identification of activating region (AR) of Escherichia coli LysR type transcription factor CysB and CysB contact site on RNA polymerase alpha subunit at the cysP promoter. *Molecular microbiology* **53**: 791-806.
10. DÌaz, E. and M. Prieto (2000). Bacterial promoters triggering biodegradation of aromatic pollutants. *Current Opinion in Biotechnology* **11**: 467-475.
11. Bundy, B., L. Collier, et al. (2002). Synergistic transcriptional activation by one regulatory protein in response to two metabolites. *Proceedings of the National Academy of Sciences* **99**: 7693.
12. McFall, S., S. Chugani, et al. (1998). Transcriptional activation of the catechol and chlorocatechol operons: variations on a theme1. *Gene* **223**: 257-267.
13. Kim, S., K. Merchant, et al. (2002). "OxyR: A Molecular Code for Redox-Related Signaling. *Cell* **109**: 383-396.
14. Porr˙a, O., M. GarcÌa Jaramillo, et al. (2007). The LysR type regulator AtzR binding site: DNA sequences involved in activation, repression and cyanuric acid dependent repositioning. *Molecular microbiology* **66**: 410-427.
15. Dubbs, P., J. Dubbs, et al. (2004). Effector-mediated interaction of CbbRI and CbbRII regulators with target sequences in Rhodobacter capsulatus. *Journal of bacteriology* **186**: 8026.
16. Van Keulen, G., A. Ridder, et al. (2003). Analysis of DNA binding and transcriptional activation by the LysR-type transcriptional regulator CbbR of Xanthobacter flavus. *Journal of bacteriology* **185**: 1245.
17. Perez-Martin, J., F. Rojo, et al. (1994). Promoters responsive to DNA

bending: a common theme in prokaryotic gene expression. *Microbiology and Molecular Biology Reviews* **58**: 268.

18. Goethals, K., Van Montagu, M. & Holsters, M. (1992). Conserved motifs in a divergent nod box of Azorhizobium caulinodans ORS571 reveal a common structure in promoters regulated by LysR-type proteins. *Proceedings of the National Academy of Sciences* **89**, 1646.
19. Harrison, S. (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**: 715-719.
20. Luscombe, N., S. Austin, et al. (2000). An overview of the structures of protein-DNA complexes. *Genome biology* **1.**
21. Huffman, J. and R. Brennan (2002). Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Current opinion in structural biology* **12**: 98-106.
22. Brennan, R. and B. Matthews (1989). The helix-turn-helix DNA binding motif. *The Journal of biological chemistry* **264**: 1903.
23. Brennan, R. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* **74**: 773.
24. Wintjens, R. and M. Rooman (1996). Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *Journal of molecular biology* **262**: 294-313.
25. Aravind, L., V. Anantharaman, et al. (2005). The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS microbiology reviews* **29**: 231-262.

26. Kwon, H., M. Bennik, et al. (2000). Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA. *Nature Structural & Molecular Biology* **7**: 424-430.

27. Rhee, S., R. Martin, et al. (1998). A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 10413.

28. Costa, R., D. Grayson, et al. (1989). Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and alpha 1-antitrypsin genes. *Molecular and cellular biology* **9**: 1415.

29. Gajiwala, K. and S. Burley (2000). Winged helix proteins. *Current opinion in structural biology* **10**: 110-116.

30. Slansky, J. and P. Farnham (1996). Introduction to the E2F family: protein structure and gene regulation. *Current topics in microbiology and immunology* **208**:1.

31. Wojciak, J., J. Iwahara, et al. (2001). The Mu repressor DNA complex contains an immobilized 'wing' within the minor groove. *Nature Structural & Molecular Biology* **8**: 84-90.

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

33. Xu, Y., R. Heath, et al. (2001). The FadR-DNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli***.** *Journal of Biological Chemistry* **276**: 17373.

34. Orth, P., D. Schnappinger, et al. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressorñoperator system. *Nature Structural & Molecular Biology* **7**: 215-219.

35. Muraoka, S., R. Okumura, et al. (2003). Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *Journal of molecular biology* **328**: 555-566.

36. Muraoka, S., R. Okumura, et al. (2003). Purification and crystallization of a LysR-type transcriptional regulator CbnR from Ralstonia eutropha NH9. *Protein and Peptide Letters* **10**: 325-329.

37. Quiocho, F. and P. Ledvina (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Molecular Microbiology* **20**: 17-25.

38. Ezezika, O., S. Haddad, et al. (2007). Oligomerization of BenM, a LysR-type transcriptional regulator: structural basis for the aggregation of proteins in this family. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* **63**: 361-368.

39. Tyrrell, R., K. Verschueren, et al. (1997). The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Structure* **5**: 1017-1032.

40. Choi, H., S. Kim, et al. (2001). Structural basis of the redox switch in the OxyR transcription factor. *Cell* **105**: 103-113.

41. Smirnova, I., C. Dian, et al. (2004). Development of a bacterial biosensor for nitrotoluenes: the crystal structure of the transcriptional regulator DntR. *Journal of molecular biology* **340**: 405-418.

42. Balcewich, M., T. Reeve, et al. (2010). Crystal structure of the AmpR effector binding domain provides insight into the molecular regulation of inducible AmpC [beta]-lactamase. *Journal of molecular biology* **400**: 998-1010.

43. Monferrer, D., T. Tralau, et al. (2010). Structural studies on the full-length LysR-type regulator TsaR from Comamonas testosteroni T-2 reveal a novel open conformation of the tetrameric LTTR fold. *Molecular Microbiology* **75**: 1199-1214.

44. Sainsbury, S., L. Lane, et al. (2009). The structure of CrgA from Neisseria meningitidis reveals a new octameric assembly state for LysR transcriptional regulators. *Nucleic Acids Research* **37**: 4545.

45. Zhou, X., Z. Lou, et al. (2010). Crystal structure of ArgP from Mycobacterium tuberculosis confirms two distinct conformations of full-length LysR transcriptional regulators and reveals its function in DNA binding and transcriptional regulation. *Journal of molecular biology* **396**: 1012-1024.

46. Ruangprasert, A. et al., Full-Length Structures of BenM and Two Variants Reveal Different Oligomerization Schemes for LysR-Type Transcriptional egulators, *J. Mol. Biol*. (2010), doi:10.1016/j.jmb.2010.09.053

47. Collier, L., G. Gaines III, et al. (1998). "Regulation of benzoate degradation in Acinetobacter sp. strain ADP1 by BenM, a LysR-type transcriptional activator." *Journal of bacteriology* **180**: 2493.

48. Ezezika, O., S. Haddad, et al. (2007). "Distinct effector-binding sites enablesynergistic transcriptional activation by BenM, a LysR-type regulator." *Journal of molecular biology* **367**: 616-629.

49. Romero-Arroyo, C., M. Schell, et al. (1995). "catM encodes a LysR-type transcriptional activator regulating catechol degradation in Acinetobacter calcoaceticus." *Journal of bacteriology* **177**: 5891.

50. Harwood, C. and R. Parales (1996). "The -ketoadipate pathway and the biology of self-identity." *Annual Reviews in Microbiology* **50**: 553-590.

51. Ezezika, O., L. Collier-Hyams, et al. (2006). "CatM regulation of the benABCDE operon: functional divergence of two LysR-type paralogs in Acinetobacter baylyi ADP1." *Applied and environmental microbiology* **72**: 1749-1758.

52. Clark, T., S. Haddad, et al. (2003). "Crystallization of the effector-binding domains of BenM and CatM, LysR-type transcriptional regulators from Acinetobacter sp. ADP1." *Acta Crystallographica Section D: Biological Crystallography* **60**: 105-108.

CHAPTER 2

The structure of BenM DNA binding domain with its cognate DNA reveals the basis for the T-N_{11}-A binding motif of LysR-Type Transcriptional Regulators[1]

Amer M. Alanazi[1], Ellen Neidle[2] and Cory Momany[1]

[1]Department of Pharmaceutical and Biomedical Sciences,

University of Georgia, Athens, GA 30602

[2]Department of Microbiology, University of Georgia, Athens, GA 30602

[1] Alanazi, A.; E.L. Neidle, and C.A. Momany. For submission to the *Proceedings of the National Academy of Sciences of the United States of America*.

ABSTRACT

BenM protein is a LysR-type transcriptional regulator (LTTR) involved in controlling the aromatic compound degradation of benzoate in the soil bacterium *Acinetobacter baylyi* strain ADP1. Atomic structures of the BenM DNA binding domain (BenM DBD) unbound and complexed with two regulatory DNA regions of its cognate promoters at *benA* and *catB* were determined by X-ray crystallography. Consistent with other LTTR structures, the BenM DNA binding domain forms a compact globular domain composed of three helices with a classic DNA recognition helix-turn-helix motif (α2-α3) with a wing-like projection between α3 and a long linker helix. BenM DBD dimers associate with two consecutive turns of the DNA helix and introduce a bend into the DNA. Two adjacent proline residues, Pro 30 and Pro 31, and Ser 33 at the N-terminal end of the recognition helix α2 create a hydrophobic pocket that interacts with the methyl group of the thymine base in the conserved DNA LTTR recognition motif T-N_{11}-A. Highly conserved Gln 29 hydrogen bonds to the first adenine base in the BenM DNA half-site recognition sequence 5'-ATAC-3', while a second protein pocket defined by Ala 28, Pro 30, and Pro 31 binds a thymine methyl group complementary to the third adenine in the recognition sequence. Arg 34 interacts with the complement of the final cytosine base. Additional amino acid side and main chain atoms interact with the phosphate backbone. The wing motif interacts mainly with the phosphate backbone of the DNA minor groove and provides structural stability that may assist in the proper positioning of the N-terminal end

of the α2 helix as well as add additional recognition specificity via minor groove A-tract recognition by Arg 53 from the wing.

Keywords: LysR-type transcriptional regulator; BenM; DNA binding domain

INTRODUCTION

LysR-type transcriptional regulators (LTTR) represent the largest group of transcriptional regulators in proteobacteria. As such, they control diverse genes involved in many cellular functions such as antibiotic resistance, CO_2 fixation, virulence factors and amino acids biosynthesis[1, 2]. While the LTTRs are one of the most common regulators, the low solubility and aggregation problems have significantly impacted and delayed structural studies of the LTTR full-length proteins. LTTRs are composed of a DNA binding domain (residues 1-67), a long linker helix (residues 70-90) and an effector binding domain (EBD) that makes up the remainder of the protein and has a structure that resembles the periplasmic binding protein fold. Small molecule regulatory ligands bind to the EBD. The effector binding domains (EBD) are more easily crystallized than the full-length proteins and numerous structures have been reported, beginning with the structures of CysB and OxyR[3, 4]. CbnR was the first LTTR protein for which a full-length structure was determined and this was followed by CrgA, ArgP, TsaR and recently BenM[5, 6, 7, 8, 9]. At this time, no complexes of the proteins with their operator-promoters have been presented, either truncated or full-length, limiting the understanding of how these proteins regulate transcription at a molecular level. Crystal structures of the full-length LTTRs have demonstrated that the DBD assumes a classic helix-turn-helix DNA binding motif with two monomers arranged with dyad symmetry. The most common DNA binding motif is composed of the sequence T…N_{11}…A, though this can vary in both base pair composition and length[10]. LTTR genes are often divergently oriented with respect

to genes that they activate and as such can act as repressors of their own transcription while acting as transcriptional activators or repressors of the remaining genes in their regulons.

Only a few well-characterized LTTRs like CysB, OxyR, and GcvA have explored how RNA polymerase activation is achieved[11, 12, 13]. So far, there appear to be several modes of recruitment of RNA polymerase to the promoter used by LTTRs. In some cases, like CysB, OxyR, and MetR, LTTRs bind to their target promoter sites upstream of -35 regions to activate the transcription via interaction of the C-terminal domain of the holoenzyme α subunit with the activator[11, 12, 14]. Mutational studies at the contact site of α-CTD showed a disruption of the interaction between α-CTD and the regulator in the DBD region in residues Y27, T28, and S29[13]. Other activation sites that might be involved in the contact site between α-CTD and activator have also been identified by mutational mapping, which include the turn between the two helices of the HTH motif and the C-terminus of the LTTR effector binding domain[13].

BenM is a well-characterized LTTR from the soil bacterium *Acinetobacter baylyi* strain ADP1[15]. In conjunction with CatM, BenM regulates a complex pathway involved in degradation of the aromatic compound benzoate to TCA intermediates via gene products encoded by the *ben* and *cat* operons (Figure. 2.1). DNAse protection analysis showed that BenM interacts with three palindromic sites (consensus 5'-A<u>T</u>AC…N_7….GT<u>A</u>T) in the *benA* promoter[16]. In the absence of effectors, BenM covers sites 1 and 3 and thereby act as a repressor of its own transcription and the *benA* promoter. When the physiological

inducers benzoate and *cis*, *cis*-muconate (muconate, hereafter) are present, the protein undergoes a conformational change where it associates with sites 1 and 2 and thereby acts as an activator of *benA* transcription. While benzoate and muconate can both bind to what is termed the primary effector binding site, which is localized at the interface of the two effector binding sub domains, benzoate is able to bind at a second site where it may work synergistically with muconate for maximal gene expression[17]. CatM, while substantially similar to BenM in amino acid sequence, does not recognize benzoate and has different promoter specificities from BenM. In the DNA binding domain, CatM and BenM are 98% identical. Despite the conserved palindromic consensus sequence shared by both BenM and CatM, BenM does not operate on the *catB* promoter[15, 17].

Here we describe the crystal structures of the unbound BenM DBD protein and the BenM DBD bound to *benA* and *catB* site 1 DNA complexes. The crystal structures of BenM DBD dimer bound to DNA reveal the molecular basis for sequence-specific recognition and binding between two consecutive turns of the DNA major groove. Because of the highly conserved nature of the DNA binding domain of LTTRs, the results apply broadly to the family and may help to explain the sequence specificity of LTTRs.

MATERIALS AND METHODS

Chemicals

Reagent grade chemicals and 18 M Ω cm^{-1} water were used. Fluka puris grade (>99.9%) imidazole used for protein purification had UV absorbent

impurities evident at 280 nm. Oligonucleotides were synthesized by IDT at 0.2 nMole scales and used without further purification.

Cloning, purification, and crystallization of BenM DBD

An expression vector (pBAC 952) encoding the DNA binding domain and linker helix of BenM, representing the first 87 amino acid residues and a C-terminal hexahistidine purification tag, was cloned by excising the DNA encoding the effector-binding domain from the full-length gene (pBAC 433). This was achieved by PCR mutagenesis using 5'-phosphorylated primers 5'GGGCACCACCACCACCACCAC (forward primer) and 5'CGAGGCAATGCGCTTGG (reverse primer internal in the *benM* gene), followed by self-ligation of the gel purified PCR product. The forward primer introduced a new glycine residue in the protein sequence before the C-terminal hexahistidine purification tag. The construct was verified by sequencing and transformed into BL21 (DE3) RIL cells (Stratagene) for protein expression. The transformed cells were grown overnight at 37 °C in 100 mL of autoinduction media[18], harvested by centrifugation at 7,000 g for 10 minutes at 4 °C and the pellets resuspended in 12 mL of binding buffer (20 mM Tris-Cl, pH 8.0, 0.5 M NaCl, 20% glycerol, 10 mM β-mercaptoethanol, 5 mM imidazole). The suspended cells were lysed using a prechilled French pressure cell at 16,000 psi. The cell lysate was then centrifuged at 39,000 g for 30 minutes at 4 °C and the supernatant was loaded onto a 1 ml HisTrap metal chelate column (GE Biosciences) charged with Ni^{+2} equilibrated in the binding buffer, and the protein

was eluted using a linear gradient of binding buffer containing 500 mM imidazole. Protein yield was 40 mg purified protein per 100 ml of cell culture. SDS-PAGE confirmed that the monomeric species was the appropriate size of 10 KDa, and gel filtration analysis showed that a dimeric species was the predominant oligomeric species. The purified BenM DBD protein was concentrated from the metal chelate fractions to 7 mg/ml using Millipore Ultrafree centrifugal concentrators at 4 °C for crystallization after dialysis with 20 mM Tris-Cl pH 8.0, 0.1 M NaCl, 10% glycerol, 10 mM BME, and 0.5 mM EDTA. The protein concentration was determined using a BioRad micro assay with BSA as the standard.

DNA complex preparation and crystallization trials

DNA duplexes were first prepared by annealing complementary oligonucleotides (0.1 mM each) in 10 mM Tris-Cl, 10 mM Nacl, 1 mM EDTA, pH 7.5 (Table 2.1). To prepare BenM DBD/DNA complexes, the BenM DBD protein was incubated with the annealed *benA* site 1 or *catB* site 1 DNA duplexes separately for 30-45 minutes on ice at the ratios of 1:1.2, protein to DNA respectively. The BenM DBD/DNA complexes were then concentrated to various degrees dependent on their solubility behavior in the ranges of 5 mg ml^{-1} to 20 mg ml^{-1}. Specifically, the BenM DBD/*catB* site 1 complex was concentrated to 20 mg ml^{-1} (protein concentration) before the crystallization-screening set-up, while the BenM DBD/*benA* site 1 was concentrated to 9.6 mg ml^{-1} (protein concentration).

Unbound BenM DBD protein was crystallized via the microbatch under paraffin oil method in house at 15 °C by screening crystallization kits from Hampton Research (Index screen I and II). Crystals that diffracted to the highest resolution (1.8 Å) were obtained by mixing 2 μl concentrated protein (in the metal chelate elution buffer) with an equal volume of 0.1 M HEPES pH 7.5, 10 % PEG 6000, 5% MPD. Crystallization trials of the DNA-protein complexes utilized the reagent specific crystallization kits from Hampton Research (ammonium sulfate, PEG 6000, sodium malonate and sodium chloride) using the microbatch under oil method at 15 °C incubation. In general, at most, 5 conditions from each kit with different concentrations around neutral pH were sampled. If a mixture precipitated immediately, the lower concentration of precipitate was used until setups were clear upon initial mixing. The BenM DBD/*catB* site 1 crystals were grown under Al's oil with the following conditions: 1) 2.4 M sodium malonate pH 6.0 2) 0.1 M MES pH 6.0, 2.4 M ammonium sulfate, 3) 0.1 M citric acid pH 5.0, 1.6 M ammonium sulfate. The BenM DBD/*benA* site 1 was crystallized only with one condition: 25 mM bis-Tris pH 6.5, 25 % PEG 3000, 50 mM ammonium acetate after screening and optimization.

X-ray data collection, structure determination and refinement of the BenM DBD and DNA complexes

Diffraction data were collected at the Advanced Photon Source, SBC-CAT beamline ID-19 and SER-CAT ID-22 in Argonne, IL from crystals pre-frozen in liquid nitrogen and shipped in a dry dewar. Data collected at the SER-CAT beamline was collected remotely using robotic pucks. The detector distances,

total oscillation range and oscillation widths were optimized to reduce overlaps at high resolution and obtain complete data sets using the HKL3000 prediction routines. Data were processed and scaled with HKL 3000 at the beamline and reprocessed with HKL2000 on home workstations[19]. The initial unbound BenM DBD structure crystallized in space group $C222_1$ and X-ray data were collected to 1.8 Å resolution. The structure determination of the unbound BenM DBD crystals involved standard molecular replacement using the DBD/linker helix of CbnR (residues 1- 100 of chain A, PDB: 1IXC) as the search model in the program AmoRe. In the DBD region, CbnR and BenM are 50% identical in sequence. A molecular replacement solution was readily found for a single monomer in the asymmetric unit. After rigid body refinement of helices within the HTH domain and the linker helix, an atomic model was built into the calculated density that revealed appropriate sequence differences between BenM and CbnR. The final model included residues 1 to 87. The C-terminal polyhistidine purification tag was not ordered and the region from 44 to 59 was poorly defined. Application of crystallographic symmetry to the subunit in the asymmetric unit created a dimeric unit similar to the CbnR DBD/LH dimer. Hydrogens were included in the refinement at the last few cycles (R_{free} dropped 2%). The final refined structure had an R_{factor} 18.56 % and R_{free} 23.21% after TLS refinement with 0.005 Å r.m.s.d. on bond lengths and 0.917° angle (Table. 2.1). Refinement of the structures was performed using REFMAC5 version 5.5.0072 with TLS domains, interspersed with cyclic model building and water identification using

COOT[20]. TLS domains were identified by the TLS server http://skuld.bmsc.washington.edu/~tlsmd/)[21].

The BenM DBD/*catB* site 1 complex crystallized in $P4_22_12$ space group and the X-ray data were collected to 3.1 Å resolution (Table 2.1). The Mathews parameter suggested that several molecules of BenM were likely to be present in the asymmetric unit, so the BenM DBD dimer structure was used as a search model using the various space group permutations possible with the *4/mmm* laue symmetry. The best molecular replacement solution positioned two dimeric units in the $P4_22_12$ space group. To speed the modeling process, several DNA structures were extracted from the protein data bank and fit via molecular replacement and then mutated to the correct sequence. We had suspected that the oligonucleotides, despite being asymmetric in sequence, would assume two random orientations with respect to the BenM DBD dimer in the crystal lattice. However, applying a two-fold rotation of the DNA double helical model, followed by REFMAC refinement, produced clearly worse statistics with residual difference density features in the DNA regions than the final refined model. Thus, there are subtle differences in the structures that are utilized in the crystal packing that result from differences in the DNA sequences. REFMAC refinement parameters for the nucleic acids were modified to match published parameters[22, 23]. The BenM DBD/*catB* site 1 complex model was refined to 18.22 % R factor and 21.53% R_{free} factor after TLS refinement with 0.012 Å r.m.s.d geometry of bond lengths and 1.379° angles. Finally, the BenM DBD/*benA* site 1 structure was solved by molecular replacement as well using the already solved BenM

DBD/*catB* site 1 structure as a search model. The structure was refined to 20.74 % R factor and 28.01 % R_{free} factor after TLS refinement with 0.011 Å r.m.s.d geometry of bond lengths and 1.666° angles (Figure 2.1). Figures were prepared using PyMol Molecular Graphics system, version 1.3, Schrödinger, LLC.

Sequence and structural alignment

A DBD multiple sequence alignment of LTTRs was prepared in the BioEdit sequence alignment program using a compilation of sequences in UNIPROT corresponding to the LTTR DBD domain.

RESULTS

Crystallization, structure determination, and structure of the unbound BenM DNA binding domain

The BenM DNA binding domain was cloned with a hexahistidine purification tag by genetically removing the EBD-encoding region from the plasmid used to express the full-length BenM his-tagged protein. Large amounts of purified BenM DBD protein, in the range of 40 mg per 100 mL of autoinduction culture medium, could be isolated from BL21 (DE3) cells. Intriguingly, parallel experiments with expression plasmids designed to produce CatM DBD protein did not produce soluble protein (unpublished results). Crystallization trials produced several sets of conditions that produced crystals. The best data set came from crystals where precipitant was 0.1 M HEPES pH 7.5, 10 % PEG 6000, 5 % MPD. However additional precipitant cocktails suitable for X-ray

crystallography included: 0.1 M HEPES pH 7.5, 8 % ethylene glycol, 10 % PEG 8000; 0.1 M Tris pH 8.5, 25 % *t*-butanol; 0.1 M Tris pH 8.5, 20 % ethanol; 0.1 M bicine pH 9.0, 0.1 M Nacl, 20 % PEG MME 550; and 0.1 M bicine pH 9.0, 2 % dioxane, 10 % PEG 20,000. Data sets were collected from crystals isolated from these different conditions, but the diffraction quality were significantly lower the best condition.

X-ray data were collected for the BenM DBD/*catB* site 1 crystals grown from 2.4 M sodium malonate pH 6.0 as the precipitant. Crystals were also obtained from precipitants containing 0.1 M MES pH 6.0, 2.4 M ammonium sulfate and 0.1 M citric acid pH 5.0, 1.6 M ammonium sulfate. The centered orthorhombic crystals of the unbound BenM DBD had a single molecule of BenM DBD in the asymmetric unit that could form a biologically relevant dimer by application of crystallographic two-fold symmetry (Figure 2.2). As one would expect from the high sequence similarity among LTTRs in the DBD, the BenM DBD structure is similar to other structures of the DNA binding domains of LTTRs as found in CbnR[5] and the later full-length structures of CrgA, TsaR, ArgP, and BenM[6, 7, 8, 9]. As such, the N-terminal helices assume a classic helix-turn-helix (HTH) motif found in the DNA binding domains of many transcriptional regulators[24, 25, 26, 27, 28, 29]. The BenM DBD consisted of α1 (residues 1-15), a turn region (residues 16-17), α2 (residues 18-25), a second turn region (residues 26-28), α3 (residues 29-43), a wing (residues 44-58), and the linker helix (residues 59-83) (Figure 2.3.). The position of the “wing” is different from the classic “winged” HTH in which the wings normally directly interrupt, precede, or follow

the HTH motif [30, 31, 32, 33, 34, 35]. As one would expect with 1.81 Å resolution data, the overall electron density of the unliganded structure was quite good. However the wing had weaker electron density with respect to the rest of the structure and less clearly defined secondary structure. This is particularly different from the well-defined CbnR DBD wing, which has two antiparallel β-sheet making a β-hairpin between the DNA recognition helix and the linker helix of the protein[5]. In the ArgP DNA-binding domain structure, the compact subunit has the wing before the linker helix, whereas the extended subunit has no antiparallel β-strands in the wing region[7]. One last difference of note is with the uncharacterized LTTR structure from *Pseudomonas aeruginosa* (PDB ID code: 2ESN), which has an extra helix of 10 residues at the N-terminus before the three helices of the DBD domain. With this particular LTTR, a N-terminal polyhistidine purification tag would project away from the protein, where it would not be likely to interfere with folding of the protein or DNA interactions (though likely increase non-specific interactions). In BenM, the amino terminal end of helix α1 faces the two-fold related α1 helix in the structure of BenM. This unfavorable dipole arrangement is stabilized by Glu 2. Met 1 starts to extend out of the helix, and many LTTRs have N-termini that would coincide with this region. Polyhistidine purification tags placed on N-termini are often used to simplify protein purifications. But N-terminal histidines will be optimally placed to interact with the phosphate backbone of DNA based on the *benA* and *catB* site 1 structures. Thus, we caution that a wide range of artifacts with regard to DNA interactions may be generated when LTTRs are used with N-terminal his tags. In

the case of full-length BenM, the protein is soluble with an N-terminal purification tag, but not functional when the gene encoding such a tag is tested *in vivo*[9].

Crystallization of BenM DBD/DNA complexes via a minimalistic screen and structure determinations of the complexes

Like many other DNA binding proteins, BenM DBD binds to a unique inverted repeat DNA sequence. In our case, we chose to crystallize oligonucleotides with authentic promoter sequences, sites 1 and 2 of the *benA* and *catB* promoters, that contain the inverted repeat sequence ATAC. Thus, the oligonucleotides were not completely symmetric. Crystallization trials of the BenM DBD/DNA complexes included minor permutations of the DNA sequences with varying overhangs and lengths for regions in the *benA* and *catB* promoters (Table 2.2). We focused particularly on the site 1 region as the half-sites were most conserved there, but screened site 2 oligonucleotides as well. A DBD dimer was anticipated to bind to two half-sites. BenM does not normally activate transcription at *catB*, so we were curious to see if there would be structural differences with between the *catB* and *benA* DNA. The largest differences in the sequences of the promoter regions are not in the site 1, but in the site 2 (Figure 2.1). The choice of the lengths of the DNA were based on earlier foot-printing studies[36], which suggested that fragments around 25 bp in size would likely be best for interaction with the DBDs.

We hypothesized that the particular oligonucleotide would be more important than the particular crystallization reagent and that current DNA

synthesis chemistries are now robust enough to produce high quality oligonucleotides in the 20 to 30 mer range. So several permutations of the lengths of the oligonucleotides were explored using unpurified oligonucleotides. The minimum commercially available scales were used for the synthesis of the oligonucleotides. The different duplex DNA was incubated with protein and a minimalistic screen of conditions favorable for nucleic acid protein complexes was manually explored. In general, at the protein concentrations used for crystallization with the minimal DNA concentrations used, only about 50 conditions could be screened. Microbatch crystallization setups were performed manually starting from conditions most likely to be productive (*i.e.* neutral pH). Only a few conditions were evaluated and the results from each individual condition were used to choose the next condition. If a precipitate resulted, a lower precipitant concentration would be used in the next well. If the well was clear, a higher precipitant concentration was used. The component specific kits used were the ammonium sulfate, malonate, and PEG 6000 screens. Only a few crystallization conditions were evaluated from each kit starting at neutral pH plus or minus one pH unit, and the condition chosen was adjusted dependent on the results of the previous mixing. The neutral pH was chosen based on the observation that most protein-DNA complexes tend to crystallize at neutral pH. If the mix of protein and precipitant showed any precipitation, we setup the lower precipitant condition. If the solution was clear, we moved to a higher precipitant concentration. In this way, only a few conditions, at most 5 per kit (out of the usual 16), were sampled from the focused screens and the conditions were

chosen based on the preceding knowledge. Using the minimal screens, crystals of the DNA complexes were obtained from as little as 50 conditions screened (i.e. a total of 100 microliters were used). While the use of robotics allows much more exhaustive screens with this amount of material, the crystals that we obtained were directly useable for data collection. We tried this strategy to crystallize BenM DBD with *benA* site 1 and site 2 separately (25 mer each), combined together (site 1 and 2, 46 mer), and *catB* site 1 and site 2 separately and combined together as well (site 1 and 2, 46 mer). We obtained crystals only with *benA* site 1 or *catB* site 1 alone.

The crystals of the protein-DNA complexes did not diffract to as high resolution as the unbound crystal. Crystals for both DNA complexes showed marked defects in the diffraction shapes, with the BenM DBD/*catB* site 1 promoter crystal being clearly cracked. Reported here are the results from the best of several data sets that were collected. The structure determinations of the DNA complexes involved molecular replacement using the BenM DBD dimeric unit as the search model. The DNA was readily visible in difference Fourier maps phased from the protein alone. Of note, the default refinement parameter set in the CCP4 suite version we used required modification with the latest nucleic acid parameters available[22, 23].

The BenM DBD/*catB* DNA site 1 complex crystallized with two dimeric units of BenM DBD in the asymmetric unit of the $P4_22_12$ crystal form with each protein monomer interacting with a single DNA half-site. Interestingly, the crystal packing of the BenM DBD subunits creates a particularly tight interaction with

face-to-face contacts reminiscent of the more hydrophobic surfaces of natural protein-protein interactions. One further oddity was an electron density feature at the N-terminal ends of the two dimers that was not easily explained with the components of the crystallization buffer (*i.e.* glycerol, EDTA, malonate, or Tris). We modeled this feature as two malonate molecules with a single sodium ion acting as a bridge between the carboxylates of the malonate. Arg 7 and Lys 83 from one chain of each dimer interact with the carboxylate ions of the malonate.

The *benA* site 1 DNA complex was crystallized in space group $P2_12_12$ with two BenM DBD dimers and two DNA duplexes as well. Surprisingly, the crystal packing of the complex was significantly different from the BenM DBD/*catB* site 1 structure. In the *benA* site 1 structure, the DNA pack end-on to form continuous super helices in the crystal along the long crystallographic *c* axis. The 5' ends of one DNA duplex associate with the 3' end and are only lacking a phosphate group from creating a truly continuous helix. Interestingly, in the *benA* site 1 structure, the interactions between subunits are via non-crystallographic two-fold symmetry, whereas in the *catB* site 1 complex structure, the interactions are via crystallographic two-fold symmetry.

General features of the structures of the BenM DBD with *benA* and *catB* site 1 DNA

In these structures, the two recognition helices of one BenM DBD dimer occupy two consecutive turns of the major groove per DNA duplex. In many other transcriptional regulators, the two recognition helices of one dimer occupy a more

continuous region of the major groove[37]. The solvent accessible surface areas between the BenM dimeric unit and the double-stranded DNA helices were 19049.5 Å^2, 19568.4 Å^2 for *catB* site 1 and *benA* site 1 respectively. Remarkably, only a few water molecules were found in these protein-DNA complex structures. Very few water molecules are involved in either the protein-protein crystal contacts or the protein-DNA contacts. Thus, indirect read of the phosphate backbone or base sequences via waters is not likely a significant factor in this LTTR.

In contrast to the unbound protein that had broken electron density for the wing that precedes the linker helix, the wing density is better defined in the DNA complexes. The region still has some modeling ambiguity in that the side chain density is not well defined in all of the subunits. Because the interaction is located at the ends of the oligonucleotides used for crystallization, there may be some loosening of the strands that may be responsible for weakening the association of this region with the DNA. The positioning of the wing is assisted by a salt bridge between Arg 50 and Glu 40 that is not observed in the unliganded structure. An additional hydrogen bond with the carbonyl oxygen of Leu 48 locks the guanidine group of Arg 50 in place. In the unbound structure, Arg 50 is exposed to the solvent and interacts with Glu 40 through the neighboring water molecules by indirect ionic interactions. Most of the full-length structures like CbnR and TsaR also lack this critical salt-bridge. Further, this wing in general does not appear to be properly positioned for DNA recognition in the full-length structures. In most of the BenM DBD subunits, the wing forms an oval shape,

rather than an extended β-hairpin, that follows along the DNA phosphate backbone. Arg 53 projects into the minor groove of the A-rich regions 5' of the ATAC and 3' of the complementary half-site. Recognition of the minor grove has been observed in multiple DNA-binding proteins[38] and is consistent with the view that the arginine is attracted to the enhanced electrostatic charge of the squeezed minor grove that results from poly A-tracts.

Exceptions to the canonical structural interactions presented above were observed in certain subunits. The wing Arg 53 was highly varied in the structures and varied from a position in minor groove of chain A, B and near Adenine 25 to a position outside the minor grove where it interacted with the phosphate of Adenine 25 in chain G and H of BenM DBD/*catB* site 1 structure. This interaction variability between the four chains is probably due to the flexibility of the DNA at the ends of the duplex as well as sequence differences. There are two conformations of Arg 4 in chain C of BenM DBD/*catB* site one structure where one is hydrogen bonded with the phosphate group of adenine 15 just like the other chains and the other one is projected away from the phosphate group. Surprisingly, Arg 34, which interacts directly with the guanine base of G17' in *catB* site 1 in all of the four monomers of the asymmetric unity, did not appear to be associated with the same guanine in chains A and C of the *benA* site 1 structure. Indeed, Arg 34 did not even contacting the DNA in chains B and D of *benA* complex. Instead, it appears that it is reaching out to the C-terminal end of the recognition helix and making hydrogen bond contacts with Gln 37. It is unclear why Arg 34 makes G17' specific interactions in chain B and D while

interacting with Gln 37 at the recognition helix C-terminal end in chain C and D. One explanation of the structural variability might be that effector binding domain is missing, which could otherwise associate with the DBD and effect interaction changes. One might even speculate that the multiple conformations might be part of a switching mechanism.

The DNA-binding surface of BenM has, as would be expected, a large number of hydrogen bond donors and positively charged amino acids, ideal for association with the negatively charged phosphodiester backbone of DNA. Two arginine (Arg 4 and Arg 34) and two lysine (Lys 20 and Lys24) side chains provide the majority of the positive charge along the DNA backbone as well as the previously mentioned wing-residues Arg 53 and to a lesser degree Arg 50. The dipole moment of helix α2 is also well aligned to provide an interaction with the DNA backbone. Additional hydrogen bonds that contact the DNA backbone are donated by the amino acid side chains of Tyr 8, Ser 17, Thr 19, Gln 29, Ser 33, Gln 35, Gln 37, and by the main chain amide nitrogens of of Phe 18 and Ala 28. The side chain of Phe 18 also makes van der Waals contacts with the deoxyribophosphate backbone as well.

The two antiparallel linker helices of the dimer display surfaces that are in general fairly hydrophobic and would normally be interacting with the effector binding domain in the full-length protein. In the *catB* structure, this face is used as a crystal contact. Intriguingly, there is a highly negative cluster of protein residues that lies at the C-terminal end of the recognition helix and is generated by the amino acids Glu 40, Glu 41, Glu 42, Glu 49 and Glu 60. The function of

this acidic cluster is not clear. It may interact with the effector binding domain or alternatively, with RNA polymerase subunits. Such a strongly negative electrostatic potential has been previously noticed on the DNA-binding surface of the Trp repressor[39]. While the particular position of these negatively charged amino acids in the protein sequences of other LTTRs are not well conserved (except for Glu 40 which forms a salt bridge with Arg 50 in a vast majority of LTTRs, see Figure 2.3) there does appear to be significant conservation of the overall charge in this region.

The two HTH domains of BenM DBD dimers are held in proximity via the two copies of the linker helix. This long helix serves as a support for the orientation and sets the spatial restraints on the two monomers for proper DNA binding. The two HTH motifs bind in tandem to the same face of the target DNA and thereby dictate the extent of the DNA bending.

The molecular basis of the T-N_{11}-A DNA motif recognized by many LTTRs

The structures of BenM DBD with site 1 DNA of the *benA* and *catB* promoters clearly explain the molecular basis of the T-N_{11}-A DNA binding motif that is prevalent among the known DNA binding sites of LTTRs. Most of the sequence-specific interactions between BenM DBD and site 1 DNA are common to both crystal structures because both site 1 DNA sequences contained two ATAC half sites. In total, 27 amino acids of BenM DBD dimer make 36 contacts with the DNA duplex (Figures 2.4, 2.5).

While the different protein chains have weakly defined density for individual amino acid residues, with a total of 8 independent molecules in the asymmetric units of the two complex structures, a clear interaction model can be derived. The BenM DBD interactions from the chain B monomer of *catB* site 1 structure with one DNA half site made of chains E and F is the most representative of the overall interaction scheme and behaved the best in refinement. But the discussion that now follows will focus on developing a consensus picture using information from all of the subunits.

The major DNA sequence-specific interactions seen in the structures are between the N-terminal residues of the recognition helix α3 and the ATAC half-site duplex. The most critical protein residues in base-specific recognition are Ala 28, Gln 29, Pro 30, Pro 31 and Arg 34, all clustered on the recognition helix α3 and all projecting into the major grove of the DNA. Working from the 5' end of the half-site (i.e. from the 5' end of *benA* promoter), the first adenine base of the ATAC sequence interacts with Gln 29 through hydrogen bonds from the Oε1 and Nε2 atoms of Gln 29 to the N6 and N7 atoms of the adenine base. The side chain of glutamine (or asparagine) is ideal for recognition of adenine due to the combination of hydrogen bond donor and acceptor that is accessible in the major grove at adenine bases. This glutamine is highly conserved in LTTRs at this amino acid position, an observation that we will return to latter in the discussion. The next nucleotide encountered, T7 (in ATAC), interacts via van der Waals interaction of its methyl group with a hydrophobic pocket defined by the side chains of Pro 30 and Ser 33, and only in some chains, Arg 34 (Figure 2.6). More

specifically, the pocket is created around the Cα, Cβ atoms of Pro 30 and the Cβ of Ser 33. After this point, the protein interactions hop to the complementary DNA chain with the thymine methyl group of the complement of A8 (AT<u>A</u>C) interacting with a second hydrophobic pocket defined by the side chains of Ala 28, Pro 30 and Pro 31, with the Cγ and Cδ atoms providing the predominant surface interactions from the prolines. The van der Waals interactions are tight around the two thymine methyl groups creating excellent complementary surfaces as shown in Figure 2.6. Finally, in most of the subunits, the guanidino side chain of Arg 34 contacts the major groove in the general region of G17', the complement of C9 (ATA<u>C</u>). The local surface of the DNA tends to have several hydrogen bond acceptors available. For instance, in several subunits of *catB* site 1 complex, the NH atoms of Arg 34 interact with the carbonyl oxygen (O6) of guanine G`17. In other subunits, the Arg 34 side chain contacts the N7 atom of A16' (*benA*) and the N7 atom of G17' (*catB*). In *benA* complex (subunits A and C), Arg 34 instead forms a salt bridge to neighboring Glu 41 and hydrogen bonds to Asn 38 and Gln 37, which contact the phosphate backbone in some subunits. In this later situation, a protein-protein crystal contact is nearby that may be responsible for changing the local electrostatics. In any case, this residue appears to have a high degree of flexibility that could be functionally important as stated previously.

Interactions from the recognition helix to the bases of the major groove ATAC sequence are clearly sequence-specific. Residues that may sense sequence-specific positioning of the phosphates via indirect read include the

previously discussed amino acids Ala 28, Ser 33, Gln 35, Gln 37, Asn 38, and Arg 50. The amide N of Ala 28 is hydrogen bonded with the phosphate group (O2P) of G17' (ATAC). The Nε2 of Gln 37 is hydrogen bonded with the phosphate group (O2P) of T7 (ATAC). The hydroxyl oxygen (Oγ) of Ser 33 and ωNH of Arg 50 are hydrogen bonded to the phosphate oxygens (O2P, O1P) of A6 (ATAC). Overall, these residues may clamp the protein against the DNA for a more sensitive read of the DNA, in particular to enhance the relatively weak van der Waals attraction associated with the two thymine methyl groups within the recognition sequence.

Sequence homology in the DBDs of LTTRs

Many of the amino acids from BenM that interact directly with the DNA are conserved among different LTTRs. Figure 2.3 shows a sequence alignment of a variety of LTTRs with a bias in our choice of sequences toward proteins that are currently well-characterized functionally or structurally. Interestingly, the strongest sequence conservation is not generally the DNA binding residues. Instead, the majority of highly conserved residues play structural roles. As an example, Ala 22 is almost absolutely conserved, but its function is to maintain the orientation of the two helices with respect to one another and is distant from the DNA. We will focus this discussion on the DNA-binding residues that are common across LTTRs.

The T-N_{11}-A motif recognition is dependent in BenM on Pro 30, Ser 33 and Arg 34. The Pro 30 is remarkably well conserved among LTTRs with only

similarly short side chain replacements like Ser, Ala, and Gly being tolerated at this position. Mutations of the residues at positions structurally equivalent to Pro 30 of BenM DBD caused drastic reductions in DNA binding in the LTTR NahR[40] and unrelated DNA binding proteins, Fis and ModE[41, 42]. Ser 33 is substantially conserved in LTTRs and plays a dual role. Ser 33 provides the surface for methyl-group recognition of the thymidine via its Cβ and also forms hydrogen bonds via Oγ to the phosphate oxygen of A6. CbnR contains a threonine at this position, which should be compatible with the dual role, and might in fact enhance binding by filling with its extra methyl group a slight void that is visible in our structures. Gln 29 is almost absolutely conserved among the LTTR amino acid sequences of diverse bacteria. Together the Gln 29 interaction and thymidine binding pockets suggest a common level of promoter specificity for most LTTR family members that would as such recognize the DNA sequence ATA. Neighboring Ala 28 is not well conserved among LTTRs, but replacement by serine and threonine residues are structurally reasonable and such sequence variations are seen in the equivalent positions of CAP and DtxR proteins[31, 43]. Moreover, addition of a hydroxyl to Ala 28 would provide enhanced interaction with the phosphate backbone. Ser 32, which is remarkably well conserved, is clearly important. Mutation of S33N in the OxyR of *E. coli* abolishes DNA binding[44]. The corresponding mutants, S34R of CysB[45] and S38P of GcvA[13], similarly caused loss of DNA binding.

According to our DNA complex structures, the wing interacts with the minor groove adjacent to the major groove of the specificity pockets. The minor

groove regions of both *benA* site 1 and *catB* site 1 have upstream A-tracts that could create narrow minor grooves optimal for electrostatic interaction of the arginine guanidinum group. Arg 53 plays this role in BenM, with assistance from Arg 50 and Glu 40, which help position the wing. Supporting the important role of Arg 50 in DNA recognition, when the corresponding amino acid is mutated to a tryptophan residue in a variant of the *E. coli* OxyR, DNA binding is abolished[44]. We anticipate that many LTTR recognition sites will include poly A runs near the recognition sequences because of the high conservation of the wing features and retention of Arg 53 in many LTTRs (though not all) or an arginine near it in the amino acid sequence (Figure 2.3). Despite the co-occurrence of Arg 50 and Glu 40 in many LTTR sequences, the amino acid sequences are quite variable among LTTRs in this region otherwise. Intriguingly, the most divergence between BenM and its paralog, CatM occurs here both in the wing itself and in residues that contact it. So the specific residues in the wing may play a subtle role in recognition specificity. In the LysR-type transcriptional regulator ArgP from *Mycobacterium tuberculosis*, strong electron density differences existed between the open and closed conformations of the DBD subunits around the residues equivalent to BenM's Arg 50 and Glu 40. In ArgP, Arg 55 pairs with Glu 45 in the "closed" conformation, but is exposed towards the solvent in the "open" conformation. Two conformations of Arg 55 in the DBDs were proposed to be due to the different chemical environments[7]. Similar distinctly different subunit conformations have also been reported in the full-length structure of CbnR[5]. Moreover, in the structure of TsaR, the electron density of the wing basic

residues Lys53 and Arg54 in the extended protomer are defined, whereas in the compact subunit they are disordered[8]. Two subunit conformations are a common feature of full-length LTTR structures, with the sole exception being CrgA, which has only one conformation of the monomer[6]. This remarkable structural difference between the unbound and bound structures of BenM DBD suggests that the wing assists in the appropriate positioning of the DNA recognition helix and consequently allows minor groove interactions between the DNA phosphate backbone and the wing residues. When coupled to conformationally different protomers, sequence dependent specificity can be achieved that in turn may be dependent on ligand binding, as proposed for in CbnR and ArgP[5, 7].

The relationship between CatM and BenM in the DBD

One of the central questions we wished to answer in this study is how CatM and BenM respond differently to promoters. As a first approximation, one would expect that the DNA binding domains would dictate promoter specificity. The amino residues that directly interact with the DNA in the BenM DBD *benA* and *catB* site 1 structures are highly conserved in CatM DBD with the exception of a few polar interactions on the phosphodiester backbone. Such residues in BenM include Thr 19 (replaced by Ser in CatM), Asn 38 (Lys in CatM), and Ser 52 (Phe in CatM). Only Lys 38 of CatM DBD seems a likely candidate for making a sequence-dependent contact as it could extend from the recognition helix to A15 of *catB* site 1- binding either to the phosphodiester backbone, or possibly extending to the local base, where it would be able to interact with the N7 of

purine bases, but not likely pyrimidine bases. Also supporting the role of Lys 38 in specificity is the close proximity of the residue to the base associated with a point mutation, T to A, at position -40 of the *benA* promoter, which renders CatM able to activate transcription of the *benABCDE* operon[36].

Intriguingly, there is a patch of amino acid differences between BenM and CatM clustered around the wing and the ends of α1 and α2. These residues include Ala 11 (Thr in CatM), Phe 18 (Ile), Leu 48 (Phe), Ser 52 (Phe), Val 55 (Ala), and Thr 57 (Val). Working together, a group of residues here could locally perturb the recognition of DNA outside the ATAC recognition motif, which would be consistent with the sequence deviations of the palindromic sequence that are at the 3' end of the recognition region that is closest to the -35 region of the promoter. The side chain of Phe 18 in particular forms a close contact with the DNA backbone outside the palindrome (5' of the ATAC), so an isoleucine (as in CatM) at this position can clearly interact to produce a sequence dependent change. While one can focus on sequence-dependent interaction differences, the RNA polymerase may be the critical player and be monitoring the subtle effects that protein has on the DNA structure. For instance, distortion of the minor groove by protein binding, and/or surface differences as a result of amino acid variations, may be dictating how RNA polymerase responds to the promoters. In this case, the promoter sequences responsible for CatM and BenM activation differences could be outside the recognition sequence.

DISCUSSION

The substantial sequence homology among LTTRs in the DBD, with an extreme case being BenM and CatM, suggests the possibility that the DNA binding domains of LTTRs may cross-recognize different promoters because of their high conservation of the DNA-binding residues. To work only in conjunction with their own promoters, additional factors have to play a role. DBD dimer-DBD dimer orientations are controlled by the effector domains in the full-length proteins, so the actual conformation of the oligomeric species will control access to the large-scale regions within the LTTR's unique promoter. In BenM, the tetrameric protein moves between sites 1 and 3 to sites 1 and 2 when inducers are present. The relative orientations of the binding sites recognized by different LTTRs can thus be quite different and provide specificity differences. However, CatM recognizes similar site arrangements as BenM. So this difference cannot provide the sole answer. Instead, additional local promoter features may be important for coupling RNA polymerase binding response to ligand-induced conformational changes of the LTTR. Access of an α-CTD from the RNA polymerase (or σ domain contacts) may require all of the following- binding of the DBD appropriately within the promoter, the presence of proper binding sequences for the α-CTD and finally, access to the site by ligand-induced changes within the effector binding domain.

Accession codes

The atomic coordinates (PDB ID codes: 3M1E, 3MY8) have been deposited in the Protein Data Bank (PDB), Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ.

ACKNOWLEDGMENTS

We would like to thank the staff at the Southeast Regional Collaborative Access team (SER-CAT) 22-ID and Structural Biology Center Collaborative Access Team (SBC-CAT) 19-ID beam lines at the Advanced Photon Source (APS), Argonne National Laboratory for their excellent assistance as well as Dr. Lirong Chen and Dr. John Rose for local support with the remote data collection. The authors would like also to thank the Saudi Arabian Cultural Mission for their sponsorship of the first author research training. The US Department of Energy, Office of Science, and Office of Basic Energy Sciences under Contract No W-31-109-Eng-38, supported use of the Advanced Photon Source. The research was funded by National Science Foundation grants MCB-0346422 (to C.M.), MCB-0516914 (to E.L.N.).

REFERENCES

1. Henikoff, S., G. Haughn, et al. (1988). "A large family of bacterial activator proteins." *Proceedings of the National Academy of Sciences of the United States of America* **85**: 6602.

2. Schell M (1993) “Molecular biology of the LysR family of transcriptional regulators”. *Annual Reviews in Microbiology* **47**: 597-626.

3. Tyrrell R*, et al.* (1997) “The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement”. *Structure* **5**:1017-1032.

4. Choi H*, et al.* (2001) “Structural basis of the redox switch in the OxyR transcription factor”. *Cell* **105**:103-113.

5. Muraoka S*, et al.* (2003) “Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend”. *Journal of molecular biology* **328**:555-566.

6. Sainsbury, S., L. Lane, et al. (2009). "The structure of CrgA from Neisseria meningitidis reveals a new octameric assembly state for LysR transcriptional.regulators." *Nucleic Acids Research* **37**: 4545

7. Zhou, X., Z. Lou, et al. (2010). "Crystal Structure of ArgP from Mycobacterium tuberculosis Confirms Two Distinct Conformations of Full-length LysR Transcriptional Regulators and Reveals Its Function in DNA Binding and Transcriptional Regulation". Journal of molecular biology **396**: 1012-1024.

8. Monferrer, D., T. Tralau, et al. (2010). "Structural studies on the full-length LysR-type regulator TsaR from Comamonas testosteroni T-2 reveal a novel open conformation of the tetrameric LTTR fold". *Molecular Microbiology* **75**: 1199-1214.

9. Ruangprasert, A. et al., "Full-Length Structures of BenM and Two Variants Reveal Different Oligomerization Schemes for LysR-Type Transcriptional Egulators." *J. Mol. Biol*. (2010), doi:10.1016/j.jmb.2010.09.053

10. Harrison S & Aggarwal A (1990) "DNA recognition by proteins with the helix-turn-helix motif". *Annual Review of Biochemistry* **59**: 933-969.

11. Lochowska, A., R. Iwanicka Nowicka, et al. (2004). "Identification of activating region (AR) of Escherichia coli LysR type transcription factor CysB and CysB contact site on RNA polymerase alpha subunit at the cysP promoter". *Molecular microbiology* **53**: 791-806.

12. Tao, K., N. Fujita, et al. (1993). "Involvement of the RNA polymerase subunit C-terminal region in co-operative interaction and transcriptional activation with OxyR protein". *Molecular microbiology* **7**: 859-864.

13. Jourdan A & Stauffer G (1998) "Mutational analysis of the transcriptional regulator GcvA: amino acids important for activation, repression, and DNA binding". *Journal of bacteriology* **180**: 4865-4871.

14. Jafri, S., M. Urbanowski, et al. (1995). "A mutation in the rpoA gene encoding the alpha subunit of RNA polymerase that affects metE-metR transcription in Escherichia coli". *Journal of bacteriology* **177**: 524.
15. Collier, L., G. Gaines III, et al. (1998). "Regulation of benzoate degradation in Acinetobacter sp. strain ADP1 by BenM, a LysR-type transcriptional Activator". *Journal of bacteriology* **180**: 2493.
16. Bundy, B., L. Collier, et al. (2002). "Synergistic transcriptional activation by one regulatory protein in response to two metabolites". *Proceedings of the National Academy of Sciences* **99**: 7693.
17. Ezezika, O., S. Haddad, et al. (2007). "Distinct effector-binding sites enable synergistic transcriptional activation by BenM, a LysR-type regulator". *Journal of molecular biology* **367**: 616-629.
18. Studier, F. (2005). "Protein production by auto-induction in high density shaking cultures." *Protein Expr. Purif* **41**: 207–234.
19. Minor, W., M. Cymborowski, et al. (2006). "HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes". *Acta Crystallographica Section D: Biological Crystallography* **62**: 859-866.
20. Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics". *Acta Crystallographica Section D: Biological Crystallography* **60**: 2126-2132.

21. Painter, J. and E. Merritt (2006). "Optimal description of a protein structure in terms of multiple groups undergoing TLS motion". *Acta Crystallographica Section D: Biological Crystallography* **62**: 439-450.
22. Gelbin, A., B. Schneider, et al. (1996). "Geometric parameters in nucleic acids: sugar and phosphate constituents". *J. Am. Chem. Soc* **118**: 519-529.
23. Clowney, L., S. Jain, et al. (1996). "Geometric parameters in nucleic acids: nitrogenous bases". *J. Am. Chem. Soc* **118**: 509-518.
24. Gajiwala K*, et al.* (2000) "Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding". *Nature* **403**: 916-921.
25. Okajima T*, et al.* (2008) "Response regulator YycF essential for bacterial growth: X-ray crystal structure of the DNA-binding domain and its PhoB-like DNA recognition motif". *FEBS letters* **582**: 3434-3438.
26. MartiAnez-Hackert E & Stock A (1997) "Structural relationships in the OmpR family of winged-helix transcription factors". *Journal of molecular biology* **269**: 301-312.
27. Wang S, Engohang-Ndong J, & Smith I (2007) "Structure of the DNA-binding domain of the response regulator PhoP from Mycobacterium tuberculosis". *Biochemistry* **46**: 14751.
28. Rhee, S., R. Martin, et al. (1998). "A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator". *Proceedings of the National Academy of Sciences of the United States of America* **95**: 10413.
29. Harrison S (1991) "A structural taxonomy of DNA-binding domains". *Nature* **353**: 715-719.

30. Huffman J & Brennan R (2002) “Prokaryotic transcription regulators: more than just the helix-turn-helix motif”. *Current opinion in structural biology* **12**: 98-106.

31. Arkinson G*, et al.* (1996) “Structure of the CAP-DNA complex at 2.5 Å resolution: a complete picture of the protein-DNA interface”. *Journal of molecular biology* **260**:395-408.

32. Ramakrishnan V (1997) “Histone structure and the organization of the nucleosome”. *Annual review of biophysics and biomolecular structure* **26**: 83-112.

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

34. Clubb R*, et al.* (1994) “A novel class of winged helix–turn–helix protein: the DNA–binding domain of Mu transposase”. *Structure* **2**: 1041-1048.

35. Harrison C, Bohm A, & Nelson H (1994) “Crystal structure of the DNA binding domain of the heat shock transcription factor”. *Science* **263**: 224-227.

36. Ezezika, O., L. Collier-Hyams, et al. (2006). "CatM regulation of the benABCDE operon: functional divergence of two LysR-type paralogs in Acinetobacter baylyi ADP1." *Applied and environmental microbiology* **72**: 1749-1758.

37. Xu Y, Heath R, Li Z, Rock C, & White S (2001) “The FadR-DNA Complex transcriptional control of fatty acid metabolism in *Escherichia coli”*. *Journal of Biological Chemistry* **276**: 17373-17379.

38. Rohs, R., S. West, et al. (2009). "The role of DNA shape in protein-DNA recognition". *Nature* **461**: 1248-1253.

39. Guenot J, Fletterick R, & Kollman P (1994) "A negative electrostatic determinant mediates the association between the Escherichia coli trp repressor and its operator DNA". *Protein Science* **3**:1276.

40. Schell M, Brown P, & Raju S (1990) "Use of saturation mutagenesis to localize probable functional domains in the NahR protein, a LysR-type transcription activator". *Journal of Biological Chemistry* **265**: 3844-3850.

41. Yuan H*, et al.* (1991) "The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding". *Proceedings of the National Academy of Sciences* **88**: 9558-9562.

42. Hall D*, et al.* (1999) "The high-resolution crystal structure of the molybdate-dependent transcriptional regulator (ModE) from Escherichia coli: a novel combination of domain folds". *The EMBO Journal* **18**: 1435.

43. Chen C, White A, Love J, Murphy J, & Ringe D (2000) "Methyl Groups of Thymine Bases Are Important for Nucleic Acid Recognition by DtxR". *Biochemistry* **39**: 10397-10407.

44. Kullik I, Toledano M, Tartaglia L, & Storz G (1995) "Mutational analysis of the redox-sensitive transcriptional regulator OxyR: regions important for oxidation and transcriptional activation". *Journal of bacteriology* **177**: 1275-1284.

45. Colyer T & Kredich N (1994) “Residue threonine-149 of the Salmonella typhimurium CysB transcription activator: mutations causing constitutive expression of positively regulated genes of the cysteine regulon”. *Molecular Microbiology* **13**: 797-805.

Table. 2.1. Crystallographic Data collection and refinement statistics.

Parameter	Unbound BenM DBD	BenM DBD/*benA* site 1 DNA	BenM DBD/*catB* site 1 DNA
Beamline[a]	19-BM	19-BM	22-ID
Wavelength (Å)	1.0080	0.97934	0.99999
Cell constants (Å)	*a* = 33.36, *b* = 84.14, *c* = 68.73	*a* = 46.00, *b* = 58.95, *c* = 300.29	*a* = *b* = 156.50, *c* = 141.53
Space group	$C222_1$	$P2_12_12$	$P4_22_12$
Resolution range (highest resolution) (Å)	50.0-1.8 (3.0 -1.8)	150-3.0 (3.5-3.0)	50.0-2.9 (3.0-2.9)
Completeness (%)	98.4	90.5	92.7
Redundancy	2.7	3.7	8.2
R_{merge} (%)[b]	4.5	6.5	11.4
Average I/σI	17.6	21.1	12.8
R_{value} (%)[c]	18.56	20.74	18.22
R_{free} (%)[d]	23.21	28.01	21.53
Number of solvent molecules Other molecules	97 1 Na^+	19 none	41 1 Na^+, 2 malonate
r.m.s.d. bond lengths (Å)	0.005	0.011	0.012
r.m.s.d. bond angles (°)	0.917	1.666	1.379
Average B-factors ($Å^2$)	23.36	70.66	93.20
Ramachandran distribution (%) Residues in most favored regions Residues in additional allowed regions	 100.00 0.00	 87.07 9.48	 95.70 2.58

[a]The Advanced Photon Source, Argonne, IL, USA, beamline 19-ID was operated by Structural Biology Consortium Collaboratory Access Team (SBCCAT) and 22-ID by the SouthEast Regional Collaboratory Access Team (SER-CAT).

[b]R_{merge} is the unweighted R value on I between symmetry related reflections.

[c]R_{value}= / for reflections in the working data set

[d]5% of the reflections were used in the cross-validation data set.

Table. 2.2. Oligonucleotide sequences used in the crystallization screening trials

DNA	Sequence[1]	Size
benA site 1	5′-TAAAAATACTCCATAGGTATTTTAT 3′-ATTTTTATGAGGTATCCATAAAATA	25
benA site 2	5′-TTATTATACAAATAATGTGTTTGAA 3′-AATAATATGTTTATTACACAAACTT	25
benA site1-site 2	5′-TAAAAATACTCCATAGGTATTTTATTATACAAATAATGTGTTTGAA 3′-ATTTTTATGAGGTATCCATAAAATAATATGTTTATTACACAAACTT	45
catB site 1	5′-TTTATATACCTTTTTAGTATGCAAA 3′-AAATATATGGAAAAATCATACGTTT	25
catB site 2	5′-CAAAAATACCAAATTGTTTATCTTT 3′-GTTTTTATGGTTTAACAAATAGAAA	25
catB site 1 short	5′-TATATACCTTTTTAGTATGCA 3′-ATATATGGAAAAATCATACGT	21
catB site 1 short overhang	5′-ATATACCTTTTTAGTATGCA 3′-TTATATGGAAAAATCATACG	20

[1] Nucleotides corresponding to the T-N_{11}-A recognition motif used by LTTRs are shown colored red. The 5'-ATAC-3' half-site sequence recognized by BenM is underlined.

FIGURE CAPTIONS

Figure 2.1. Aromatic compound degradation in *Acinetobacter baylyi* ADP1. The top line shows the metabolic intermediates as they are processed by the enzymes regulated by the *ben* and *cat* operons. Benzoate and muconate act as regulator molecules for BenM. The second line shows the gene organization of the *ben* and *cat* operon. BenM predominatly controls expression of the *benABCD* regulon. At the *catA* promoter, BenM plays an equally important regulatory role as its transcriptional paralog, CatM protein. BenM does not activate transcription of the *catB* promoter. Two DNA sequences corresponding to the *benA* and *catB* promoters are shown below the regulon. BenM and CatM recognize the half-site motif 5' ATAC.

Figure.2.2. The structure of a BenM DNA binding domain dimer. The DBD monomer consists of three helices α1, α2, and α3, followed by a "wing" that precedes a long linker helix. The structure contains a classic helix turn helix motif for DNA recognition with α3 being the DNA recognition helix. The asymmetric unit has only one BenM DBD monomer in the unbound structure. The other subunit of the homodimer has been generated by crystallographic symmetry.

Figure. 2.3. Multiple sequence alignment with secondary structure assignment. The highly conserved residues among the LTTRs are highlighted in black and gray colors with the three helices of the domain labeled on the top sequence

along with the linker helix. Critical residues discussed in the text are denoted. The sequence alignment was created with the program Bioedit using output from CLUSTALX. The protein names, organisms and SwissProt sequence identifiers are: BenM, *Acinetobacter baylyi* ADP1, O68014; CatM, A*cinetobacter baylyi* ADP1, PO774; CbnR, *Ralstonia eutropha,* Q9WXC7; DntR, *Burkholderia* sp DNT, Q7WT50; OxyR, *E. coli*, P0ACQ4; CysB, *Salmonella typhimurium*, P06614; TsaR, *Comamonas testosteroni* T-2, P94678**;** CrgA, *Neisseria meningitides*, Q9JPU9; ArgP, *Mycobacterium tuberculosis*, P67665; GcvA, *E. coli*, P0A9F6; AmpR, *Enterobacter cloacae*, P05051; MetR, *Salmonella typhimurium,* P0A2Q4; LysR, *E. coli*, P03030; NahR, *Pseudomonas putida,* P10183; NodD, *Rhizobium meliloti,* P03031; TrpI, *Pseudomonas aeruginosa,* P11720; IciA, *E. coli* K12, P0A8S1.

Figure. 2.4. BenM DBD interactions with *catB* site 1 DNA.
Shown are residues at the N-terminal end of the recognition helix $\alpha 3$ and the wing of BenM DBD monomer (chain B) that interacting with the DNA in the *catB* complex. Hydrogen bond interactions are represented as red dotted lines. The DNA chains E and F are colored magenta and yellow respectively.

Figure. 2.5. Summary of amino acid-nucleotide interactions.
The DNA sequence shown running up/down represents the DNA sequence of *benA* site 1. Nucleotide sequences in *catB* that differ form *benA* are shown within parentheses. The two ATAC….GTAT half site motifs are colored**.** The arrows

from amino acid residues to the nucleotide label denote sequence specific contacts with the base of the nucleotide**,** arrows to the deoxyribose sugar denote interactions with the deoxyribose sugar and arrows to the phosphates denote interactions with the phosphate backbones.

Figure. 2.6. The T- N_{11}-A specificity pockets.

Amino acids are shown as CPK spheres with van der Waals radii. Dots represent the van der Waals surfaces of the nucleotides. The methyl groups of the thymine bases pack tightly against residues from the recognition helix.

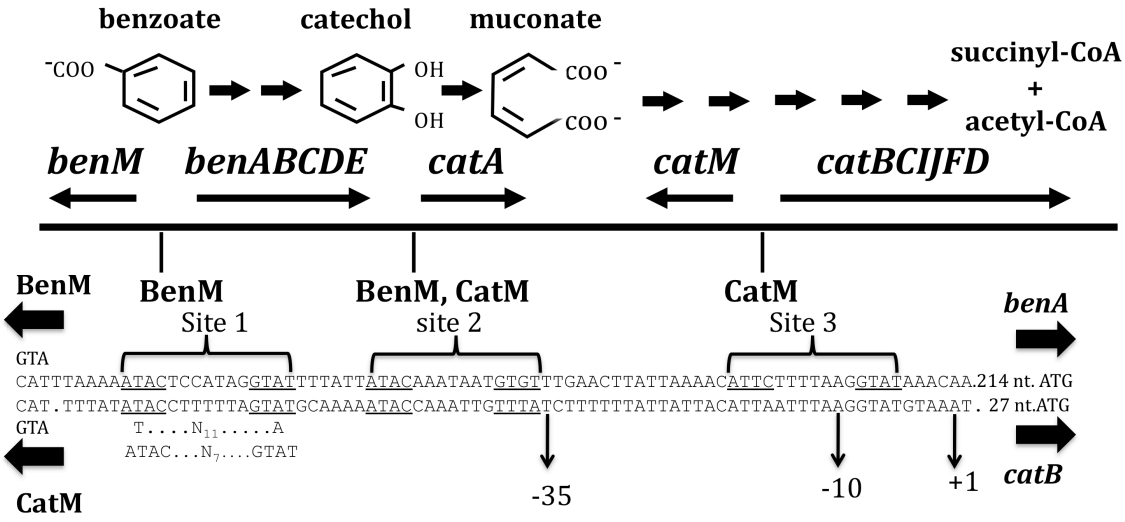


Figure 2.1. Aromatic compound degradation in *Acinetobacter baylyi* ADP1.

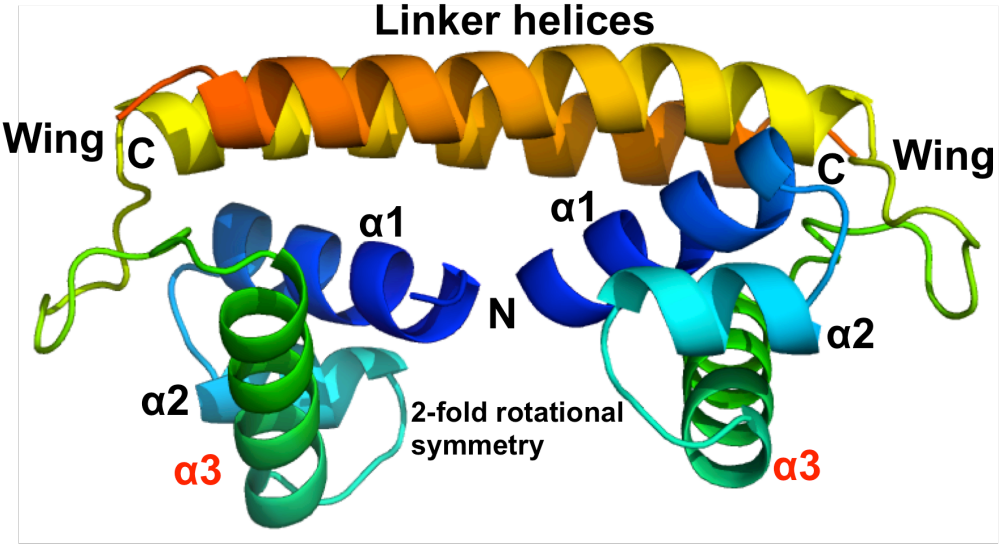


Figure.2.2. The structure of a BenM DNA binding domain dimer.

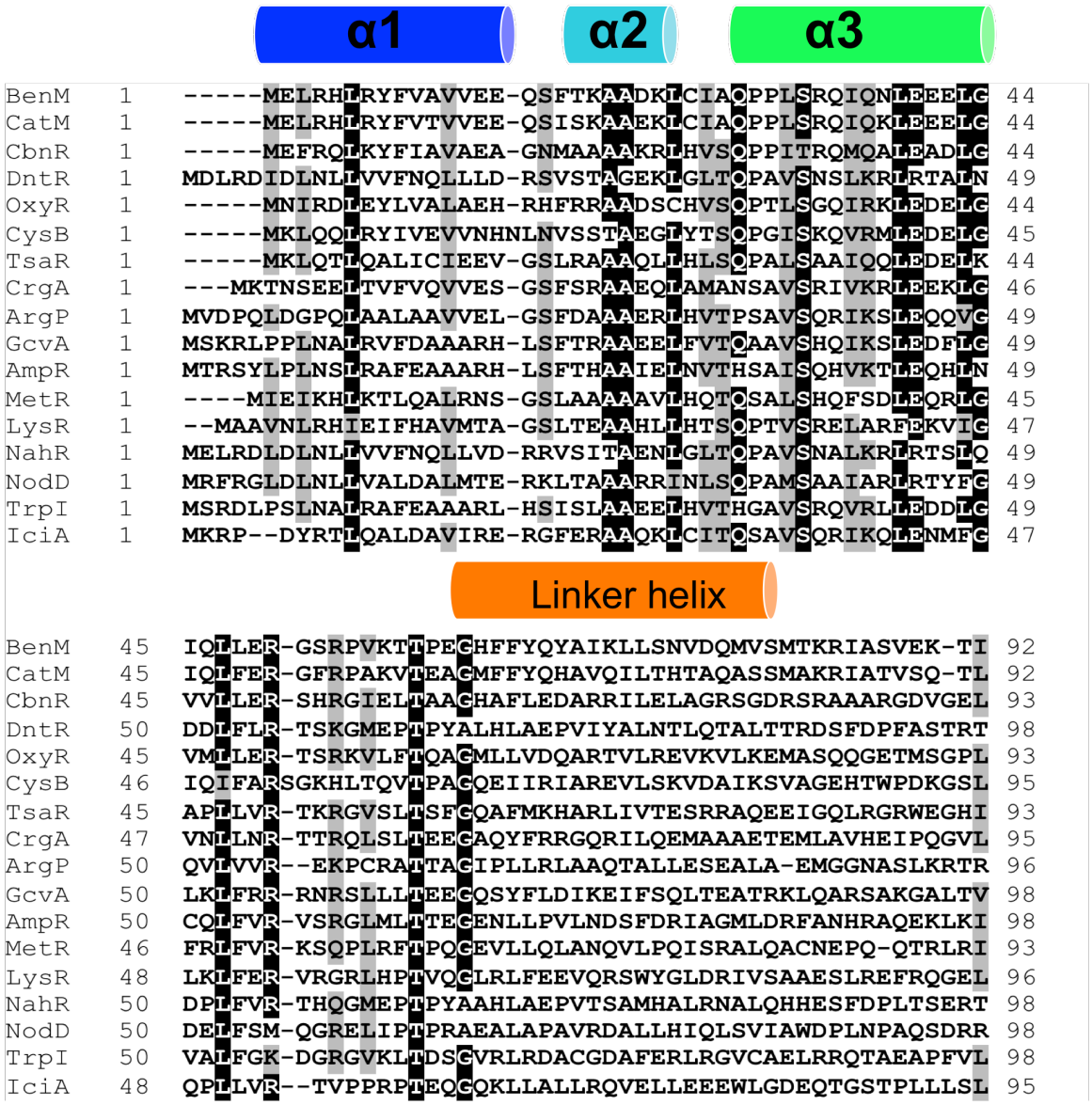


Figure. 2.3. Multiple sequence alignment with secondary structure assignment.

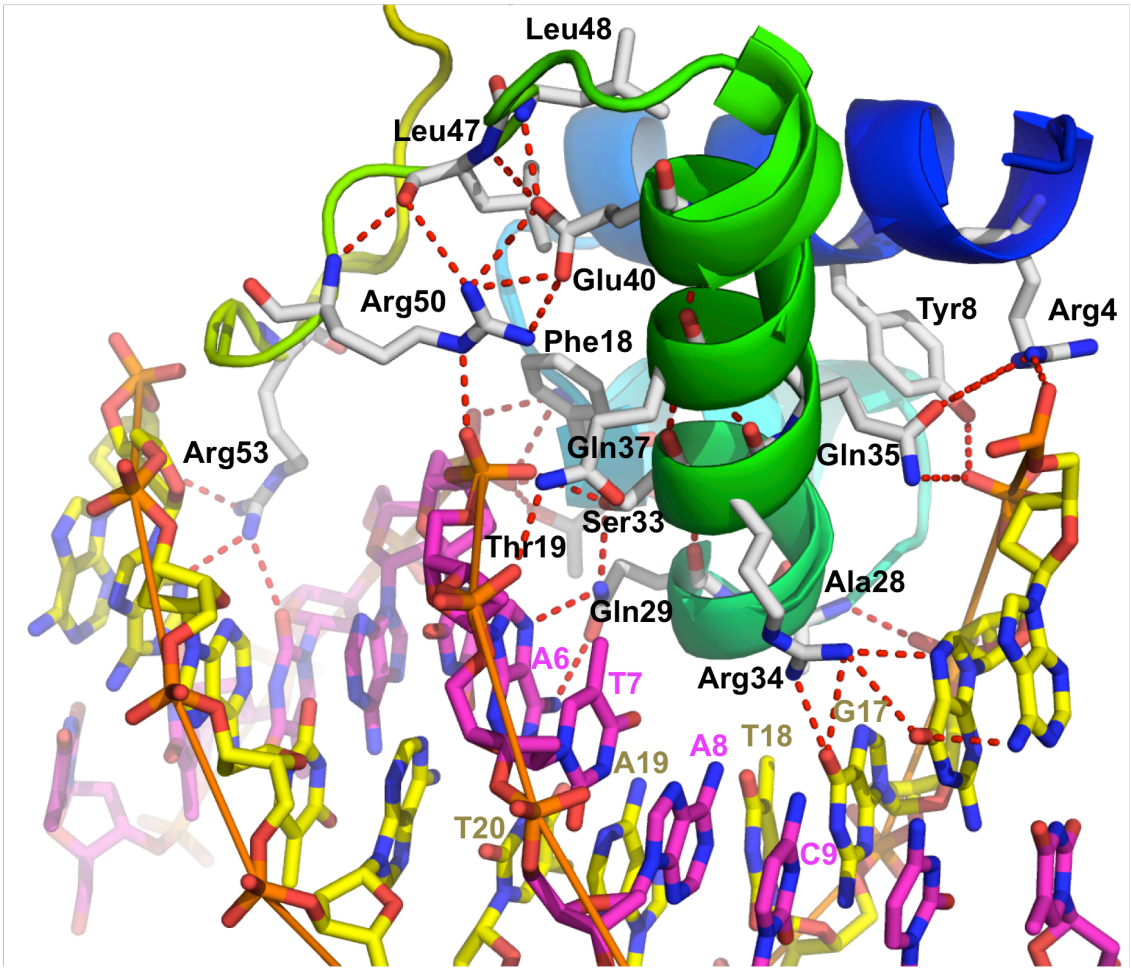


Figure. 2.4. BenM DBD interactions with *catB* site 1 DNA.

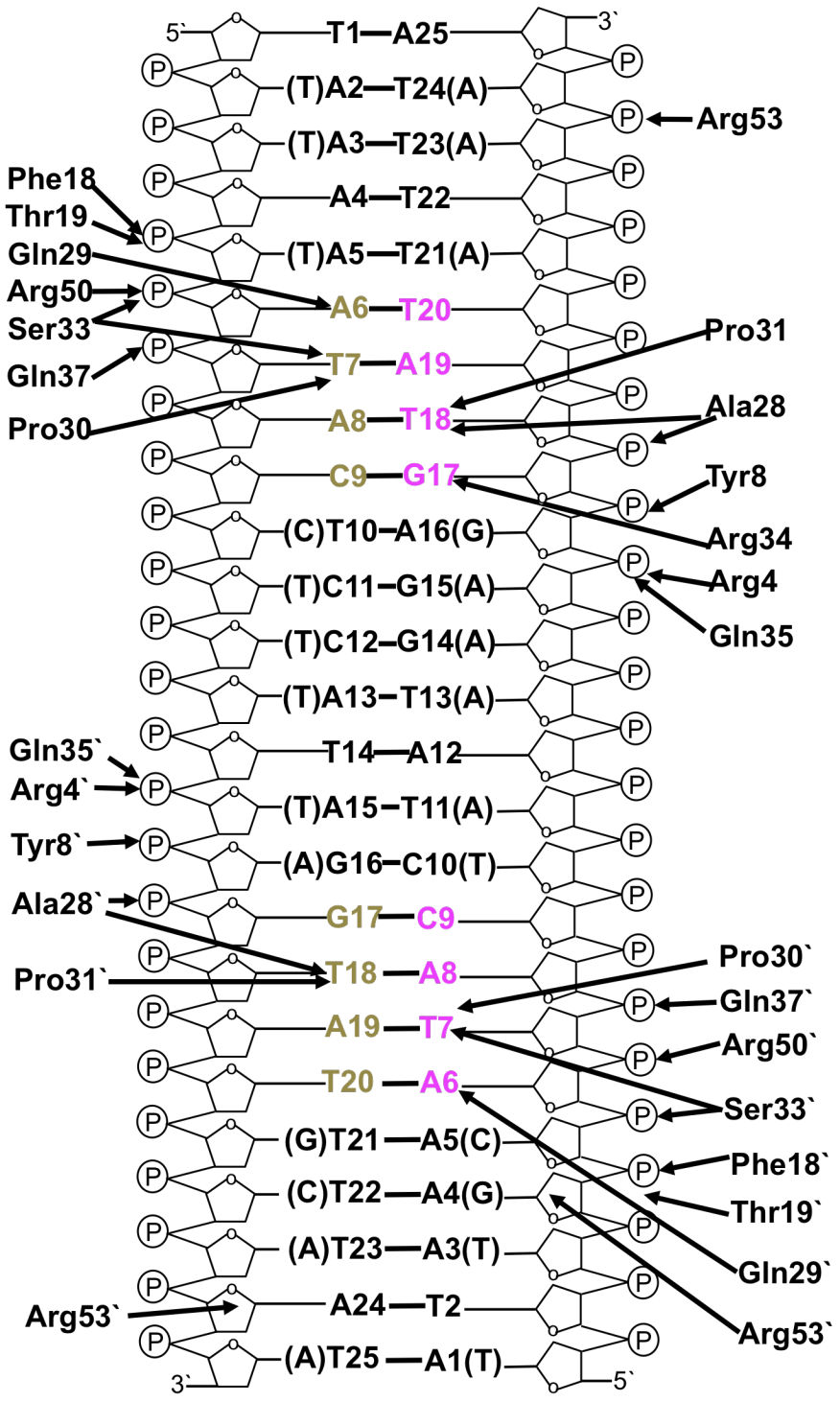


Figure. 2.5. Summary of amino acid-nucleotide interactions.

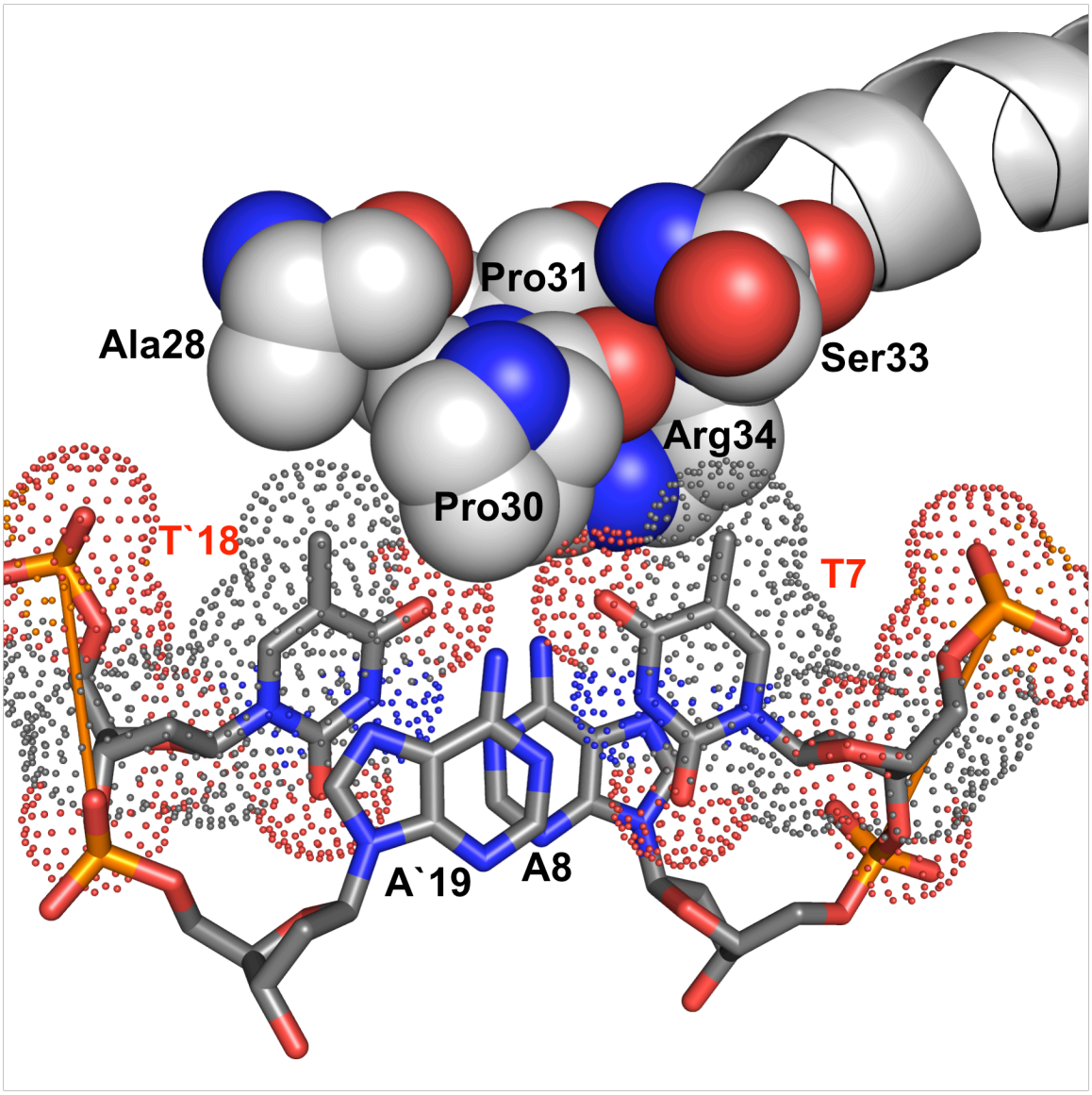


Figure. 2.6. The T- N_{11}-A specificity pockets.

CHAPTER 3

CRYSTAL PACKING OF THE LYSR-TYPE TRANSCRIPTIONAL REGULATOR BENM DNA BINDING DOMAIN WITH ITS COGNATE *CATB* SITE 1 DNA GENERATES A UNIQUE AND STABLE THREE DIMENSIONAL CRYSTAL LATTICE[1]

[1] Alanazi, A.; E.L. Neidle, and C.A. Momany. For submission to *Acta crystallographica section F: Structure communication*.

Abstract

BenM protein is a LysR-type transcriptional regulator found in *A. baylyi* ADP1. LysR-type transcriptional regulators (LTTR) are the largest family of transcriptional regulators in proteobacteria. The solubility problems and aggregation tendency of the full-length LTTR proteins slowed down many crystallization and structural studies of both the unbound protein and the target DNA promoter-bound protein. Based on the first LTTR-DNA complex structure: BenM DNA binding domain (BenM DBD) with *catB* site 1 DNA, we found many unique structural features and interactions between the protein chains and the DNA. The structure crystallized in a tetragonal crystal lattice ($P4_22_12$) and showed a tight and yet interesting packing between the asymmetric unit molecules. The asymmetric unit of the complex contains two BenM DBD dimers where each dimer binds to one *catB* site 1 DNA duplex. The oligonucleotide size being used for crystallization was critical for the proper positioning of the dimeric species and allows for the packing and binding of the domain. A molecule of malonate, used as the precipitating agent in the crystallization of the complex, was found to stabilize the dimer crystal packing and allow for the unique association between the linker helices of chains A and C. The duplex DNA made of chains G and H packs end-to-end to create a continuous helix with the other symmetry related DNA duplexes and runs perpendicular to the crystallographic *c* cell axis. The DNA chains E and F instead run as two columns parallel to the crystallographic *c* cell axis with the symmetry related mates in a unique protein-dependent fashion. The dimer-dimer interface contains mostly ionic residues. All

of these unique and interesting crystal packing features along with the use of the specific precipitating agent might be used to engineer a crystal lattice for studies of other LTTR-DNA complexes.

Index words: BenM; DNA binding domain, Lys-R-type transcriptional activator, crystal engineering.

Introduction

Protein-protein crystal contacts are usually random, non-specific but complex and precisely balanced. These contacts have been neglected owing to the observation that proteins adopt wide polymorphism. The crystal packing contacts are different from the natural (*i.e.* physiological) protein-protein contacts, which play an important role in cell recognition and signal transduction (Durbin *et al.*, 1996, Carugo *et at*. 1997). Association of protein molecules through weak forces forms protein crystals and the packing can differ drastically depending on the neighbouring environment of the protein molecules. Therefore, the use of salts and precipitating agents all can change the nature of the crystal packing leading to different crystal forms while the protein structure remains the same in all forms. Depending on the protein solution, the entire protein surface can be involved in the contacts as in the case of the pancreatic ribonuclease enzyme, which has been crystallized in different crystal forms, but all of them have the same protein structure (Crosio *et at* 1992).

The LysR-type transcriptional regulator (LTTR) BenM protein controls the genes responsible about benzoate degradation in the soil bacterium *Acinetobacter baylyi* ADP1. LTTRs are the largest family of regulators in proteobacteria that control genes and regulons of diverse cellular functions such as virulence factors, $C0_2$ fixation and antibiotic resistance (Schell, 1993; Henikoff *et al.*, 1988; Pareja *et al.*, 2006). Based on mutational analysis, the N-terminus of the protein involved in the DNA binding which extends from residue 1 to approximately 68 amino acids. This region (i.e. DNA binding domain) has been

associated with the insolubility and aggregation problems seen with the previous LTTR full-length structural studies. As a result of that, many truncated versions of LTTR have been crystallized where the DNA binding domain (DBD) have been removed and the solved structure represent the effectors binding domain (EBD) of the protein (Verschueren *et al.*, 1999; Tyrrell *et al.*, 1997; Choi *et al.*, 2001; Stec *et al.*, 2006). Another mechanism that might contribute to the solubility issues with this family of proteins is the formation of high order oligomers as it has been found with two BenM EBD variants and demonstrated that oligomerization might be a common feature among LTTR (Ezezika *et at*. 2007).

We analyzed here the crystal packing of BenM DBD protein (residues 1~89) with *catB* site 1 DNA (25 mer) complex structure from *A. baylyi* ADP1. This DNA site represents one of the DNA promoter sites that have been confirmed by DNase foot printing for BenM protein where it contains the well-known LTTR recognition motif: T-N_{11}-A (Bundy *et al.*, 2002, figure 3.1). The complex crystal is unique in its crystal packing, where it forms a three dimensional arrangement and orientation of the DBD monomers in the asymmetric unit and the DNA serves as a cross bridging between the DBD monomers. We detailed many interesting interactions and crystal packing features between the protein monomers and the DNA that can explain that unique crystal packing and the two-fold symmetry of the complex structure.

Materials and methods

Plasmid construction, expression, and purification of BenM DBD

An expression construct for his-tagged BenM DBD was created by PCR amplifying plasmid pBAC430, which encodes the *Acinetobacter baylyi* strain ADP1 full length BenM protein in a pET21b-based vector (Sambrook E.J.F *et al*, 1989), using 5'-phosphorylated primers 5'GGGCACCACCACCACCACCAC (forward primer) and 5'CGAGGCAATGCGCTTGG (reverse primer), followed by self-ligation of the gel purified PCR product. The forward primer introduced a new glycine residue in the protein sequence before the six-histidine residues of the C-terminal polyhistidine purification tag. The construct was verified by sequencing and transformed into BL21 (DE3) RIL cells (Stratagene) for protein expression. The transformed cells were grown overnight in 100 mL of auto induction media (Studier F, 2005), harvested by centrifugation at 7,000 g for 10 minutes at 4 °C and the pellets resuspended in 12 mL of the purification binding buffer (20 mM Tris-cl, pH 8.0, 0.5 M NaCl, 20% glycerol, 10 mM β-mercaptoethanol, 5 mM imidazole). The suspended cells were lysed using a prechilled French press cell at 1,000 psi. The cell lysate was then centrifuged at 39,000 g for 30 minutes at 4 °C. The supernatant was loaded onto a 1 ml Ni^{+} metal chelate column (GE Biosciences) equilibrated in the binding buffer and eluted using a linear gradient of binding buffer containing 500 mM imidazole. The purified protein fractions were evaluated by Tris-tricine SDS-PAGE and the pooled fractions dialyzed into 20 mM Tris-Cl pH 8.0, 0.1 M NaCl, 10% glycerol and 10 mM β-mercaptoethanol. The purified BenM DBD protein was concentrated to 7 mg/ml using a Millipore

ultra free concentrators at 4 °C for crystallization. The protein finally was quantified by Bio-Rad micro assay.

BenM DBD and *catB* site 1 complex preparation and crystallization

The DNA was first annealed (0.05 mM) in 10 mM Tris-Cl, 10 mM NaCl, 1 mM EDTA, pH 7.5. To prepare BenM DBD/*catB* site 1 complex, the BenM DBD protein was incubated with the annealed *catB* site 1 duplex for 30-45 minutes on ice at the ratios of 1:1.2, protein to DNA respectively. The complex was then concentrated to 20 mg ml^{-1} before the crystallization-screening set-up. Crystallization trials of the complex were performed and identified by using crystallization kits from Hampton Research and the microbatch under oil method at 15 °C incubation. The BenM DBD/*catB* site 1 crystals were grown under Al's oil with the following Conditions: 1) 2.4 M sodium malonate pH 6.0 2) 0.1 M MES pH 6.0, 2.4 M ammonium sulfate, 3) 0.1 M citric acid pH 5.0, 1.6 M ammonium sulfate.

X-ray data collection, structure determination and refinement of BenM DBD/*catB* site 1 complex

Diffraction data were collected at the Advanced Photon Source, SER-CAT ID-22 in Argonne, IL from crystals pre-frozen in liquid nitrogen and shipped in a dry dewar. Data were collected remotely using robotic pucks. The detector distances, total oscillation range and oscillation widths were optimized to reduce overlaps at high resolution and obtain complete data sets using the HKL3000 prediction routines. Data were processed and scaled with HKL2000 on home

workstations (Minor, W., M. Cymborowski, *et al.* 2006). Refinement of the structures was performed using REFMAC5 version 5.5.0072 with TLS domains interspersed with cyclic model building and water identification using COOT (Emsley, P. and K. Cpwtan, 2004). TLS domains were identified by the TLS server (http://skuld.bmsc.washington.edu/~tlsmd/) (Painter, J. and E. Merritt, 2006). The complex structure was determined using first, the already solved unbound BenM DBD structure as the molecular replacement model for the BenM DBD/*catB* site 1 structure. Phases from the protein alone returned clear density corresponding to the DNA duplexes, however, molecular replacement with the DNA part of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA (PDB ID 1U78) was used as a search model for the DNA part and to position DNA fragments of the complex. The orientations of the DNA duplex was swapped and refined to confirm that the correct sequence of the DNA was assigned.

Results and discussion

Crystallization and structure determination

The BenM DBD/*catB* site 1 complex crystallized in $P4_22_12$ space group and the X-ray data were refined to 3.1Å resolution (table 1). The asymmetric unit of the complex contains two BenM DBD dimers, each dimer bound to an inverted repeat of 25 base pair of *catB* site 1 DNA that encompassed the BenM DNA recognition motif (ATAC···N_7···GTAC). Therefore, each BenM DBD monomer interacts with a single DNA half site (Figure 1). The difference Fourier maps for the DNA part of the crystal unit was clearly visible and phased out from the

protein alone. The modelling process of the structure were expedited by identifying the DNA and fit via molecular replacement with rigid body replacement.

We were assuming that there would be two random orientations of *catB* site 1 despite of its asymmetry in sequence, with respect to the lattice. However, to test that, we reversed the DNA sequences and refined them immediately with REFMAC. Therefore, worse statistics with residual difference density features in the DNA regions were produced. Based on that, there are subtle differences in the structures that are utilized in the crystal packing that result from differences in the sequences. The BenM DBD/*catB* site 1 complex was refined to 18 % R factor and 21 % R free factor after TLS refinement with 0.011Å r.m.s.d geometry of bond lengths.

Crystal packing of the asymmetric unit

Interestingly, the crystal packing of the two BenM DBD dimers creates a particularly tight interaction with face-to-face contacts reminiscent of the more hydrophobic surfaces of natural protein-protein interactions. Only a few water molecules were found in the complex *in toto* and those few water molecules were located far away from the dimer-dimer surfaces. Therefore, we were not surprised when we calculated the solvent accessible surface area and the crystal contact area of the complex with the CryCo program (Eyal *et al*, 2005) and found to be 19049.5 $Å^2$ and 561.5 $Å^2$ respectively. The total solvent accessibilities of the atoms making crystal contacts were used to calculate the solvent accessible

surface area with the symmetry related molecules apparently being removed. In fact, only a few water molecules were found bridging at the DNA-protein interface, so the complex is rather unique in that very few water molecules are involved in either the protein-protein crystal contacts or the protein-DNA contacts.

There are only 18 BenM DBD residues from three chains in the asymmetric unit, which are making contacts with the other symmetry related mates (table 2). We were expecting that the linker helices could create stable helix bundles, but the packing is far more complex with the DBD surfaces wrapping around the linker helices where linker helices of chains A and C cross one-another and the linker helices B and D run parallel to A and C respectively.

Owing to the 2-fold symmetry nature of the DBD dimers, the linker helices of each dimer are packed against each other but in the reverse direction. Therefore, the linker helix C-terminus of chain A is packed against the linker helix N-terminus of chain B and visa verse. Likewise, the linker helix C-terminus of chain C is packed against the linker helix N-terminus of chain D and visa verse. These linker helices packing are important to allow for the dimer interface interactions between the two monomer N-terminal ends in each dimer by the proper positioning of the α1 in each monomer which will be discussed later (Figure. 3.1).

In addition to the protein dimer-dimer contacts that are created by the crystallographic two-fold symmetry, the *catB* site 1 complex utilized several modes of packing in the crystal lattice that were selective for the crystallized DNA size and length. We have tried three permutations of *catB* site 1 DNA to facilitate

the DNA binding and form the complex before crystallization set-up. Those three permutations were the native 25 mer of *catB*, short *catB* (21 mer) and short overhang *catB* (20 mer). The short permutation of *catB* site 1 is missing two thymine and two adenine bases from the native site at 5` and 3` respectively. Whereas, the short overhang permutation is missing three thymine bases at the 5` end and two adenine bases at the 3` end. The thymine base overhang of that permutation has been introduced at the 3` end of the complementary strand. These permutations were synthesized to investigate which one of those different DNA sizes might crystallize better with BenM DBD. All of these permutations have been tried with the same conditions and only the native long *catB* (25 mer) gave crystals with BenM DBD. Based on the crystal structure and asymmetric unit packing, it is clear that the extra base pairs on 5` and 3` ends of each duplex was essential for the wing contact where they form a minor groove specific for the wing Arg residue binding. End-on association of the double helices is a common crystal-packing motif and this was observed in the *catB* complex.

The double helices defined by chains G and H and the crystallographic symmetry-related mates create continuous, end-to-end, helices that run perpendicular to the crystallographic *c* cell axis and parallel to the *a* and *b* axes as a result of the 4_2 screw of the crystallographic symmetry. The 5' end of chain G (of H) continues to the 3' end of a symmetry related chain G (of H) and the distances between the C5' atom to it's symmetry related C3' atom distances are 4.8 and 5.0 Å respectively. The double helices defined by chains E and F and their crystallographic symmetry related mates do not extend across the crystals

like chains G and H. Instead, they run as two columns parallel to the crystallographic *c* cell axis in an end-on fashion where, the 5' end of the E chain interacts only with the 3' end of the F chain, in a two-fold relationship. Thus, the chain does not extend throughout the crystal. Instead, the 3' end of the E chain and 5' end of the F chain terminate by packing against symmetry related protein chain A (Lys 56 with the phosphate group between residues 24 and 25 of chain E) and chain D (Arg 84 to the phosphate group between residues 1 and 2 of chain F). Further crystal packing is generated by the protein chains C and D linker helices where they run parallel to each other but in the reverse direction and pack against the 3` end of chain E and 5` end of chain F of the 2-fold DNA symmetry related mates respectively. The final crystal packing contact involves interactions of the linker helix C-terminus of the symmetry related B chain with a3 and parts of the wing between chain A linker helix and a3 (Figure. 3.2).

Malonate stabilization of the complex structure

BenM DBD/*catB* site 1 crystals were grown with high concentration of sodium malonate (2.4 M). Thus, during the model building process, a high electron density feature was clearly visible at the N-terminal ends of the two dimers that was not easily explained with the components of the crystallization buffer (*i.e.* glycerol, EDTA, or Tris). When we noticed that there are a numerous positively charged residues at the *catB* site 1 dimer-dimer interface and also close to that high electron density feature, we modelled that region as two malonate molecules (along with a single sodium ion) for being negatively

charged molecules and might stabilize that interface region. Specifically, the carboxylate ions of the malonate molecules interacting with that sodium ion and acting as a bridge between them. Owing to the linker helices crossing between chain A and C, Arg 7 inα1 and the linker helix Lys 83 from chain A of one dimer were able to charge stabilize the other carboxylate ions of one malonate residue. The other carboxylate ions of the second malonate residues were charge stabilized by the same residues but in chain C. Thus, if our interpretation of the density is correct, the malonate plays a unique role in stabilizing the BenM DBD/*catB* site 1 complex crystal lattice and perhaps assists in the cross linking between the two dimer linker helices. The specific interactions that are taking place at that region are as follow: NH1 of Arg7 in chain C and NZ of Lys83 in chain A are interacting with the O6 and O7 of malonate 1 carboxylate ion respectively. The other interactions are the NH1 and NH2 of Arg7 in chain A with the O8 and O6 of malonate 2 carboxylate ion respectively. The last interaction is between Lys83 in chain C with O9 of that latter malonate carboxylate ion (Figure. 3.3). This unique crystal lattice that might be stabilized by malonate residues is another example of crystallization engineering where the use of malonate in our complex was essential for that unique cross bridging of DNA chains and domain wrapping around themselves.

Engineering crystallizability has been tried frequently and there are many protein structures have been solved by different kinds of engineering approaches. Among those examples is the use of small lipid molecules such as cholesterol to engineer a human β_2-adrenergic G protein-coupled receptor

(Cherezov *et al.*, 2007). Cholesterol, essential component in the crystallization, was used to mediate and stabilize the parallel association of the receptor molecules in the crystal lattice. Moreover, The T4 lysozyme crystals were grown by engineering intramolecular disulfide cross-linking (Heinz *et al.*, 1994), Human H ferritin crystals grown by engineering intramolecular crystal contacts (Lawson *et al.*, 1991) and lastly the human thymidylate synthase crystals were grown by mutations of surface residues (McElroy *et al.*, 1992).

Monomer-monomer interface interactions

The two dimer interface regions in the asymmetric unit of BenM DBD/*catB* site 1 complex structure are mostly comprise of scattered positively charged residues represented by Arg4, His5 and Arg7 in each monomer. Since that dimer interface region is mainly positively charged in nature, we found that those residues are participating in an interaction with each other in a way to bring those N-terminal ends of the dimer close to each other allowing for both interaction with the negatively charged DNA phosphate backbone of the minor groove and maintaining the 2-fold symmetry dimer interface. Moreover, the minor groove area at the dimer interface between the two DNA half sites have five thymine nucleotides in a row or T-trac (five Adenine nucleotides in the complementary strand, A-trac) which make it a highly enriched region and specific for arginine binding (Rohs *et at*. 2009).

The domain and subunit interfaces have been analyzed and found that they tend to prefer an aromatic residues (Phe, Trp, Tyr, His) as well as Met and Leu over the interdomain cavities which tend to have molar polar residues.

Particularly, Histidine residue was found to be strongly favored at the domain interfaces over the other bulky aromatic residues (Hubbard *et al.* 1994, Argos *et al* 1988). In accordance with that analysis of the domain and subunit interfaces, the dimer interface of BenM DBD contains two Leu residues and one His residue along with other charged residues, which will be mentioned below. The interactions at the dimer interface are as follow: the oxygen (O) of Met1 in chain B interacting with the NH2 of Arg 7 in the other chain (chain A), the carbonyl oxygen (OE1) of Glu2 in chain B interacting with the N of Arg4 in chain A. Alternatively, the oxygen of Met1 in chain A interacting with the NH2 of Arg7 in chain B, the carbonyl oxygen (OE1) of Glu2 in chain A interacting with the N of Arg4 in chain B. The specific minor groove interactions with the N-terminal end of Arg4 residues in each monomer are the NH1 of Arg4 in chain B with the phosphate oxygen (OP1) of adenine 15 in chain F and the NH1 of Arg4 in chain A with the phosphate oxygen (OP1) of thymine 15 in chain E (Figure. 3.4). All of these interactions have also been found in the other dimer of the asymmetric unit (dimer chains C, D with DNA chains G and H).

Figure captions

Figure 3.1. BenM DBD/*catB* site 1 asymmetric unit.

The four domain chains colored as follow: chain A: green, chain B: blue, chain C: red, chain D: yellow. The DNA chains colored as follow: chain E: green, chain F: cyan, chain G: pink, chain H: yellow. The crystallized oligonucleotide sequence is labelled under the structure. Hence that the LTTR recognition motif is labelled above the sequence and the two palindromic DNA half sites are underlined. All the figures were prepared in PyMol Molecular Graphics system, version 1.3, Schrödinger, LLC.

Figure 3.2. Crystal packing of BenM DBD/*catB* site 1 complex.

Construction of the symmetry related molecules in one unit cell with the use of the CryCo (crystal contact analysis) program (Eyal *et al.*, 2005). All the chains colored as the previous figure.

Figure 3.3. Sodium malonate interactions with BenM DNA binding domain/*catB* site 1 complex.

The DBD chains are colored as in the first figure. The red dotted lines denote the ionic interactions between malonate molecules, sodium ion and the two arginine, lysine residue in chains A and C.

Figure 3.4. Dimer-interface interactions with the minor groove.
The N-terminus interactions of BenM DBD at the dimer interface and the minor groove of *catB* site 1. The red dotted lines denote the hydrogen bonding interactions between each monomer residues along with the minor groove. All the chains are colored as the first figure.

Conclusions

According to the crystal packing arrangement and features found in the first LTTR-DNA complex structure along with the DBD linker helices crossing facilitated by two malonate residues as a crystallization agent, we can conclude that the protein oligomer species play a central role in the crystallization process. In our complex, it is clearly that the dimer nature of the domain is important for the tight packing of the asymmetric unit and consequently for binding with the DNA target site. The presented crystal packing here could be utilize to engineer a crystal lattice for crystallization of other LTTR DNA complexes and guide future crystal growth studies. Since crystallization of macromolecules still considered to be an empirical process and depends entirely on the trial and error where one would screen many conditions to find crystals, the necessity to engineer a crystal lattice by taking advantage of the crystal packing and acquiring knowledge about the protein surface and domain interfaces are becoming essential.

Acknowledgments

We would like to thank the staff at the Southeast Regional Collaborative Access team (SER-CAT) 22-ID at the Advanced Photon Source (APS), Argonne National Laboratory for their excellent assistance as well as Dr. Lirong Chen and

Dr. John Rose for local support with the remote data collection. The authors would like also to thank the Saudi Arabia Cultural Mission for their sponsorship of the first author research training. The US Department of Energy, Office of Science, and Office of Basic Energy Sciences under Contract No supported use of the Advanced Photon Source. W-31-109-Eng-38. The research was funded by National Science Foundation grants MCB-0346422 (to C.M.) and MCB-0516914 (to E.L.N.).

References

Argos, P. (1988). *Protein Eng.* **2**, 101.

Carugo, O. and P. Argos (1997). *Protein Sci.* **6**, 2261–2263.

Cherezov, V. D. & Rosenbaum (2007). *Science*, **318**, 1258.

Choi, H., Kim, S., Mukhopadhyay, P., Cho, S., Woo, J., Storz, G. & Ryu, S. (2001). *Cell*, **105**, 103–113.

Bundy, B., L. Collier, et al. (2002). *Proceedings of the National Academy of Sciences* **99**: 7693.

Collier, L. G., Gaines, III, *et al.* (1998). *J. Bacteriol.* **180**, 2493.

Crosio, M. J. & Janin (1992). *J. Mol. Biol.* **228**, 243–251.

Durbin, S. and G. Feher (1996). *Annu. Rev. Phys. Chem.* **47**, 171–204.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Eyal, E. S. & Gerzon (2005). *J. Mol. Biol.* **351**, 431–442.

Ezezika, O. S. & Haddad (2007). *Acta Crystallographica Section F*: Structural Biology and Crystallization Communications **63**: 361-368.

Heinz, D. and B. Matthews (1994). *Protein Eng.* **7**, 301.

Henikoff, S., Haughn, G. W., Calvo, J. M. & Wallace, J. C. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 6602–6606.

Hubbard, S. and P. Argos (1994). *Protein Sci.* **3**, 2194–2206.

Lawson, D. P. & Artymiuk (1991). *Nature*, **349**, 541.

McElroy, H. G. & Sisson (1992). *J. Cryst. Growth*, **122**, 265–272.

Minor, W. M. & Cymborowski (2006). *Acta Crystallographica Section D*: Biological Crystallography **62**: 859-866.

Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.

Pareja, E. P. & Pareja-Tobes (2006). *BMC Microbiol.* **6**, 29.

Rohs, R. S. & West (2009). *Nature*, **461**, 1248–1253.

Sambrook J, E. F. Fritsch, and T. Maniatis (1989). Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory Press*

Schell, M. (1993). *Annual Reviews in Microbiology* **47**: 597-626.

Stec, E., Witkowska-Zimny, M., Hryniewicz, M. M., Neumann, P., Wilkinson, A. J., Brzozowski, A. M., Verma, C. S., Zaim, J., Wysocki, S. & Bujacz, G. D. (2006). *J. Mol. Biol.* **364**, 309–322.

Studier, F. (2005). *Protein Expr. Purif.* **41**, 207–234.

Tyrrell, R., Verschueren, K. H., Dodson, E. J., Murshudov, G. N., Addy, C. & Wilkinson, A. J. (1997). *Structure*, **5**, 1017–1032.

Verschueren, K. H. G., Tyrrell, R., Murshudov, G. N., Dodson, E. J. & Wilkinson, A. J. (1999). *Acta Cryst.* D**55**, 369–378.

Table 3.1. Crystallographic Data collection and refinement statistics.

Parameter	BenM DBD/*catB* site 1
Beamline[a]	22-ID
Wavelength (Å)	0.99999
Cell constants (Å)	a = b = 156.501, c = 141.538
Space group	$P4_22_12$
Resolution range (highest resolution) (Å)	50.00-2.91 (3.00-2.91)
Completeness (%)	92.7
Redundancy	8.200
R_{merge} (%)[b]	11.4
Average I/σI	12.8
R_{value} (%)[c]	18.22
R_{free} (%)[d]	21.53
Number of solvent molecules Other molecules	41 1 Na^+, 2 malonate
r.m.s.d. bond lengths (Å)	0.012
r.m.s.d. bond angles (°)	1.379
Average B-factors (Å^2)	93.20
Ramachandran distribution Residues in most favored regions (%) Residues in additional allowed regions (%)	 95.70 2.58

[a] At the Advanced Photon Source, Argonne, IL, USA, beamline 22-ID was operated by the Southeast Regional Collaboratory Access Team (SER-CAT).

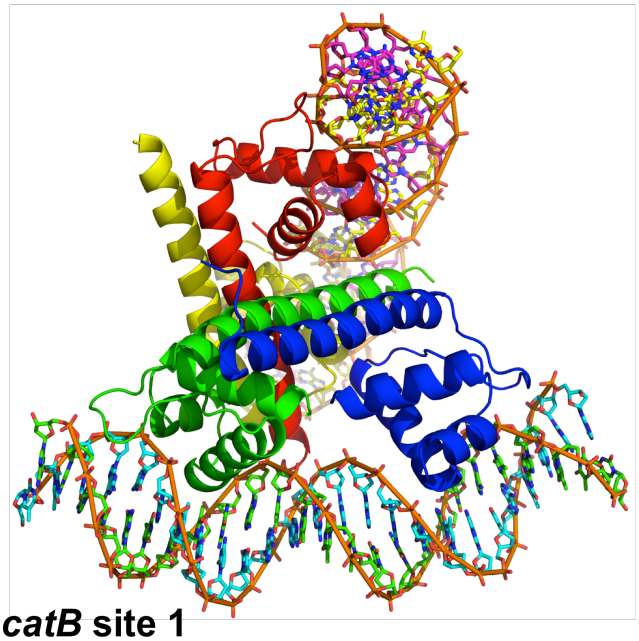
[b] R_{merge} is the unweighted R value on I between symmetry related reflections.

[c] R_{value}= Σhkl|Fobs(hkl)-Fcalc(hkl)|/ Σhkl Fobs(hkl) for reflections in the working data set.

[d] 5% of the reflections was used in the cross-validation data set.

Table 3.2. Residues participating in the crystal contacts of BenM DBD/*catB* site 1 structure along with the calculated surface area and minimal distance between each residue and the symmetry related mates. Calculations of the crystal contact surface areas and the distances were generated in CryCo (crystal contact analysis) program (Eyal *et at*, 2005) with the distance threshold set to be between 4 to 6 Å.

Residue	Crystal contact	
	Area, Å^2	Minimal distance, Å
Glu 40 A	13.4	3.2
Glu 41 A	34.4	3.6
Glu 42 A	15.8	3.7
Leu 43 A	5.8	3.1
Gly 44 A	53.2	3.5
Ile 45 A	5.8	4.0
Gln 46 A	25.6	3.5
Asp 23 B	31.9	2.8
Lys 24 B	40.0	4.4
Cys 26 B	13.0	5.2
Asp 76 B	3.0	5.4
Ser 80 B	1.1	5.1
Lys 83 B	53.7	2.9
Arg 84 B	27.7	4.4
Ala 86 B	10.1	3.5
Ser 87 B	36.6	3.6
His 89 B	27.9	3.6
Arg 84 D	18.1	4.4



***catB* site 1**

1 T. . . . N_{11} . . . . A 25

TTTATATACCTTTTTAGTATGCAAA

25 1

AAATATATGGAAAAATCATACGTTT

Figure 3.1. BenM DBD/*catB* site 1 asymmetric unit.

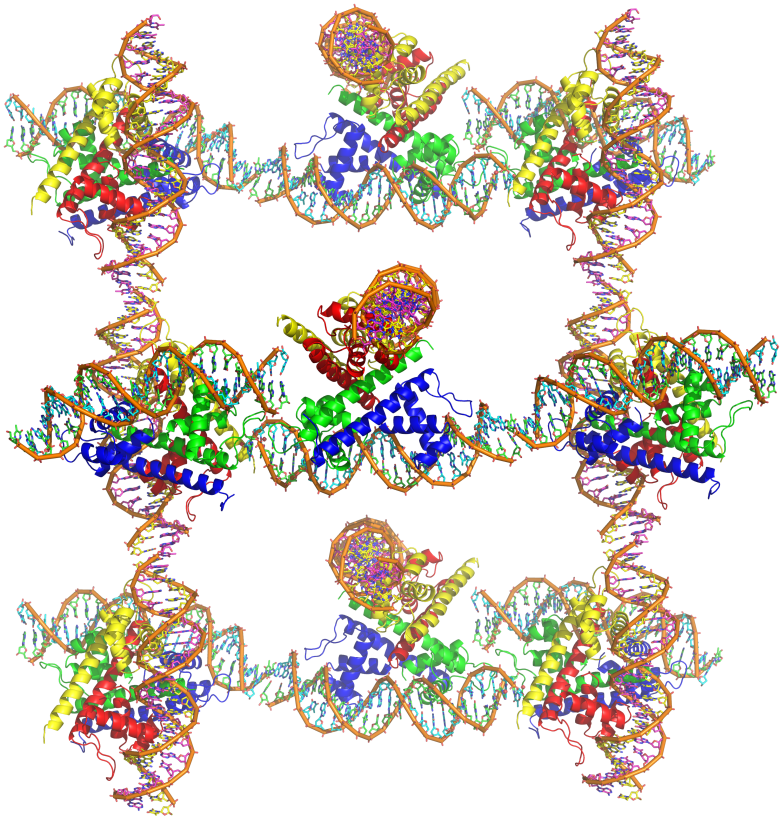


Figure 3.2. Crystal packing of BenM DBD/*catB* site 1 complex.

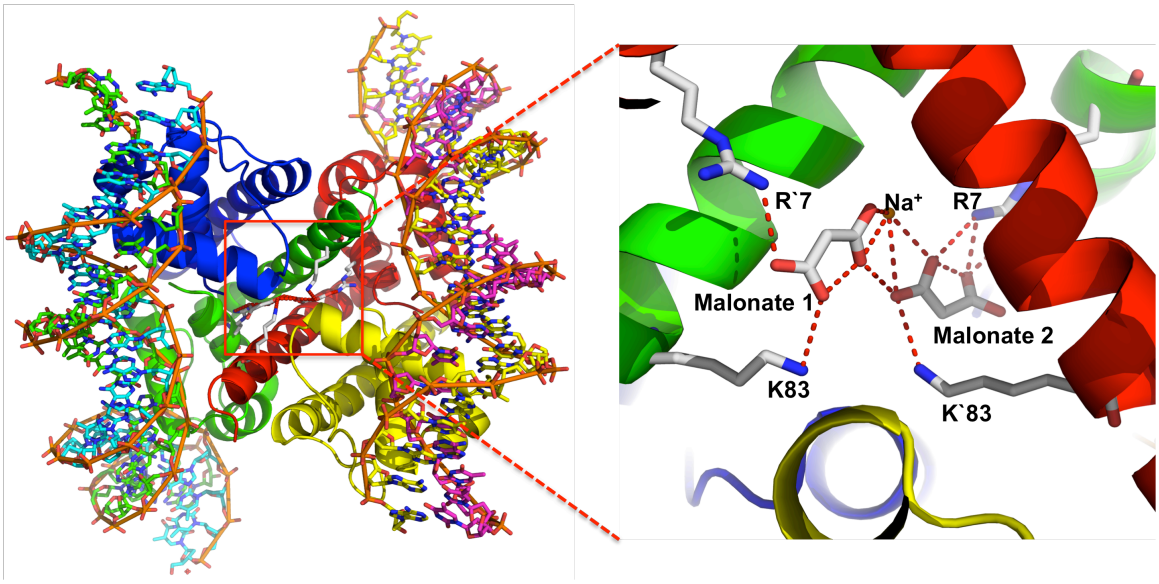


Figure 3.3. Sodium malonate interactions with BenM DNA binding domain/*catB* site 1 complex.

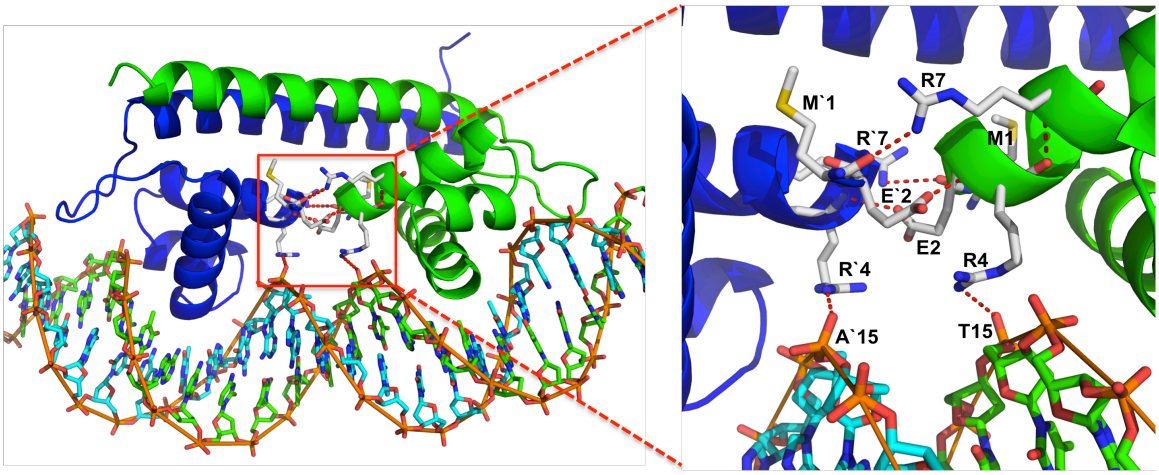


Figure 3.4. Dimer-interface interactions with the minor groove.

CHAPTER 4

CONCLUSION

The research presented in chapter 2 of this dissertation explained for the first time the molecular basis for LTTR interactions with their DNA binding sites. The atomic structures of BenM DNA-binding domain (DBD), unbound and bound with two different DNA regulatory regions, *benA* site 1 and *catB* site 1, were helpful in understanding the remarkable structural differences and the specific interactions used by LTTRs in recognizing DNA. There are currently no other LTTR DBD structures alone or with their DNA binding sites. The specific interactions of the domain with the well-known recognition motif of LTTRs, T-N_{11}-A, have been determined. The unbound BenM DBD crystallized as a monomer for which a 2-fold symmetry monomer can be generated to form the biologically active dimer. Each of the two DNA complexes crystallized as two dimers where each dimer binds to either a *benA* or *catB* site 1 duplex. However, the asymmetric unit of each complex is remarkably different.

Multiple sequence alignment of diverse LTTR DBDs showed many highly conserved residues scattered around the DBD region with the most located at the recognition helix of the HTH motif. We were able to define the specific interactions between those highly conserved residues with the T-N_{11}-A major groove and minor groove bases. The LTTR recognition specificity arises from the interaction of Gln 29 at the N-terminal end of the recognition helix with the

adenine base of the ATAC half site of regulatory region 1. Additional specificity comes from the van der Waals contacts of two small pockets created by two hydrophobic residues, Pro30 and Pro 31 at the N-terminal end of the recognition helix with the thymine base methyl groups of T-N_{11}-A motif. Interactions of Arg50, and Arg53 in the flexible coil region between the HTH motif and the linker helix with the minor groove and the C-terminal part of the recognition helix were also very important to help stabilize the domain on DNA properly.

Finally, there was a remarkable difference of the unbound and bound structures between Arg50 and Glu40 at the C-terminal end of the recognition helix. Those two residues are highly conserved among other LTTR DBDs and make direct ionic interactions in the bound complexes. In the unbound structure, the interaction was mediated via neighboring water molecules. Since the DNA regulatory region site 1 of *benA* and *catB* have the same palindromic sequence, we found the same interactions of the protein with the major and minor grooves and very minor differences in individual subunits. A minimalistic crystallization screen was very successful in identifying crystallization conditions of both DNA complexes. The protein-DNA complex structures presented here should be of great value to guide future crystallization and structural studies of LTTR-DNA complexes.

The crystallized BenM DBD/*catB* site 1 showed many interesting features in terms of crystal packing, which we discussed in more detail in chapter 3. Crystal contacts are not biologically relevant and usually random due to the isomorphic nature of protein crystals. Sodium malonate was used at high

concentration (2.4 M) in crystallizing that complex and was found to be associated with two arginine and two lysine residues each from one monomer of a different dimer. This arrangement allows for association of two linker helices from chains A and C. This lattice stabilization by malonate might be important for the three dimensional assembly of the crystal lattice. The crystal packing of two DNA duplexes within the asymmetric unit of BenM DBD/*catB* site 1 was unique. This crystal lattice might be used to engineer a different piece of DNA to crystallize other protein-DNA complexes. The other DNA complex (BenM DBD/*benA* site 1) showed different crystal packing while maintaining the same DBD interaction with each half site of *catB* site 1.