ABSTRACT

BenM and CatM are two LysR-type transcriptional regulators that activate the \textit{ben} and \textit{cat} genes needed for benzoate degradation in the soil bacterium, \textit{Acinetobacter baylyi} ADP1. Both proteins respond to a metabolite of benzoate degradation, \textit{cis,cis}-muconate. In addition, BenM, but not CatM, responds to benzoate as an effector. BenM plays the primary role in expression of the \textit{benABCDE} genes responsible for the initial catabolism of benzoate. CatM regulates the expression of the \textit{catBCIJFD} genes needed for the subsequent degradation of \textit{cis,cis}-muconate to tricarboxylic acid cycle intermediates. In the absence of BenM, CatM is not sufficient to activate the \textit{benABCDE} operon to high enough levels to enable growth on benzoate. The regulation of CatM at the \textit{benABCDE} promoter was studied in light of two mutations in CatM that enabled high level activation of \textit{benABCDE} operon, CatM(V158M) and CatM(R156H). These substitutions affected \textit{benA} and \textit{catB} gene expression differently, thus highlighting the importance of balanced expression from multiple promoters in the benzoate degradation pathway. The ability of these CatM variants to substitute for BenM underscored the evolutionary relationship between both proteins. To understand further the differences between these proteins, X-ray diffraction methods were used to solve the structures of their effector binding domains with and without their effectors, muconate and benzoate bound to the protein. The structures highlighted the important features of how both proteins respond to effectors to initiate transcriptional activation and represent the first structures of the family of LysR-type transcriptional regulators with bound effectors. These investigations postulate a molecular basis for the synergism exhibited by BenM in response to both benzoate and \textit{cis,cis}-muconate in which a charge relay system underlies the synergistic transcriptional activation. In light of these studies, the structures of the effector binding domains of CatM(V158M), CatM(R156H) and another BenM variant, BenM(R156H/T157S) were determined. In comparing these variant structures with their wild-type counterpart structures, the role that these substitutions play in affecting transcriptional regulation was addressed. In separate studies, two independent crystal structures of the BenM-effector binding domain generated from high pH conditions show how certain conditions favor oligomerization of the full-length protein that lead to solubility problems. A detailed model by which BenM, CatM and other LysR-type transcriptional regulators could form these oligomers is presented.
INDEX WORDS: *Acinetobacter baylyi* ADP1, BenM, CatM, LysR-type transcriptional regulator, benzoate, *cis,cis*-muconate, β-ketoadipate pathway, oligomerization, protein-protein interaction, synergistic induction, X-ray diffraction, crystal structure, aromatic compound degradation, transcriptional regulation, effector binding site.
STRUCTURAL AND FUNCTIONAL DIFFERENCES BETWEEN BENM AND CATM, TWO LYSR-TYPE PARALOGS IN ACINETOBACTER BAYLYI ADP1

by

OBIDIMMA CHUKWUNONSO EZEZIKA

B.Sc., THE UNIVERSITY OF LAGOS, NIGERIA, 2000

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

2006
STRUCTURAL AND FUNCTIONAL DIFFERENCES BETWEEN BENM AND CATM, TWO LYSR-TYPE PARALAGOS IN ACINETOBACTER BAYLI ADP1

by

OBIDIMMA CHUKWUNONSO EZEZIKA

Major Professor: Ellen L. Neidle
Committee: Cory Momany
Anna Karls
Robert Maier
Lawrence Shimkets

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2006
DEDICATION

I dedicate this project to my friend, hero, savior and knight in shining armor,

Jesus Christ - in whom I live, move and have my being.

"In war, then, let your great object be victory, not lengthy campaigns."
– Sun Tzu 500 BC
ACKNOWLEDGEMENTS

Every project requires the toil, time and devotion of extraordinary people. Words are not adequate to explain my gratefulness to those that have helped me with these studies. Though I fear the blunder of leaving someone out, I do want to acknowledge with great thanksgiving the following people:

- My family, especially my dad, mum and brother, Toby - whose prayers, time, encouragement, support and heartwarming conversations have been highly treasured and cherished. All this would have been impossible and futile without you. Your love and prayers have indeed brought me through this far - Thanks a million!!!
- Ellen Neidle, for shaping me as a scientist through her attention to detail and enthusiastic mentoring. Thanks a lot for always being there for me - You are truly inspiring. Your friendship and guidance are higly cherished and appreciated far more than you can ever imagine.
- Cory Momany for introducing me into the exciting world of crystallography and structural biology. Thanks a million for your time, patience and creative guidance.
- Anna Karls for her continuous stream of encouragement and guidance throughout the pursuit of these studies. Thank you!
- Rob Maier and Larry Shimkets for enlightening discussions and being great sources of inspiration. Every conversation with you two fires me up to get research done.
- Todd Clark, Sandra Haddad, Matthew Eby, Drew Reams, Becky Bundy and Lauren Collier-Hyams for making my transition into the Neidle lab a confident and smooth
one and laying the foundation from which many of the projects sprang forth - Thanks a lot!

• Sarah Craven, Hailey Dale, Andrew Burk and Ann Onyenwoke for working with me on the many different and exciting projects. Thanks for bringing out the best in me.

• Santiago Lima, for the very engaging discussions on crystallography and the multiple trips and sleepless nights collecting X-ray data - You were a remarkable colleague to work with - Thank you.

• Tim Hoover and Mark Schell for their enlightening and engaging discussions on many different aspects of this project.

• Peter Horanyi, John Rose, William Lanzilotta and James Liu for very enlightening discussions, assistance and tutoring in crystallography. You guys are spectacular!

• Benjamin Hasselbring, Lara Pereira, Jeffrey Bose, Yuchen Liu, Ashwini Nadkarni, Catherine Pennington, Jennifer Morgan, Michelle Momany, Mohammed Salameh, Todd Smith and Mella Yohannes for their support throughout these studies.

• Dale Boyer for going the extra mile with assistance in setting up many of the computer programs used in this study - Thanks.

• The staff at the high-throughput crystallization lab of the Hauptman-Woodward medical research institute for countless high-throughput protein crystallizations screens - it was definitely worth it.

• Circle of friends and relatives whose support, prayers and encouragement were highly valuable in the completion of this project - you know yourselves - Steve and Chinwe Nwajagu, the Ezenekwe family, Augustine and Kate Ogbunugwu, Laurie Fowler,
Will and Beverly Thixton, Olatoye Oyelowo, Mel and Ellen Holmes, Anthony and Jesse Costa, Steve and Allison Arnold, the McClesky family…

- Thank you all for being there....

- Finally, the Holy Spirit, whose quiet direction, guidance, inspiration and motivation were seminal in the completion of these studies - Thanks a million!
TABLE OF CONTENTS

ACKNOWLEDGEMENTS.................................................................................................................. vii
LIST OF TABLES............................................................................................................................ xii
LIST OF FIGURES ........................................................................................................................... xiii
CHAPTER

1 INTRODUCTION AND LITERATURE REVIEW ................................................................. 1
   Purpose of study ......................................................................................................................... 1
   Acinetobacter baylyi strain ADP1: History and general characteristics ......................... 3
   Aromatic compound degradation by Acinetobacter ........................................................... 5
   Regulation of genes needed for benzoate degradation......................................................... 8
   Transcriptional repression by BenM and CatM ................................................................. 9
   Muconate levels in the cell affect CatM-mediated regulation ........................................... 10
   Structure and function of LTTRs ....................................................................................... 11
   The structure of CbnR ........................................................................................................ 16
   LTTRs involved in aromatic compound degradation...................................................... 23
   Interactions of LTTRs with ligands .................................................................................... 24
   Structural characterization of the effector binding site of DntR .................................... 25
   Structures of BenM and CatM ......................................................................................... 31
   References ......................................................................................................................... 33
APPENDIX: LIST OF CRYSTALS STRUCTURES FROM THESE STUDIES

SUBMITTED TO THE PROTEIN DATA BANK WITH THEIR PDB IDENTITY

CODES .................................................................................................................................................228
LIST OF TABLES

Table 2.1: *Acinetobacter baylyi* strains and plasmids..............................................................53

Table 2.2: Effect of *catM* and *benA* mutations on rates of BenM-independent growth on benzoate..................................................................................................................................................70

Table 3.1: Crystallographic data collection and refinement statistics .................................................105

Table 4.1: List of plasmids...........................................................................................................147

Table 4.2: Crystallographic data collection and refinement statistics .................................................150

Table 4.3: Superposition of monomers of BenM-EBD, BenM-muconate-EBD and BenM(R156H/T1157S)-EBD structures ........................................................................177

Table 4.4: Superimposition of various monomers of CatM-EBD, BenM-EBD and CatM-EBD variant structures .........................................................................................................................178

Table 4.5: Average B-values of residues in β4 of wild-type and variant structures of CatM- and BenM-EBD ........................................................................................................................................179

Table 5.1: Crystal properties, data collection and refinement statistics ...........................................195

Table 5.2: Analysis of the tetrameric interface of two BenM-EBD structures with the structure of CbnR from *Ralstonia eutropha* .................................................................205

Table 5.3: Ionic and hydrogen bond interactions at tetrameric interfaces of BenM-EBD ........209
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The $\beta$-ketoacid pathway in strain ADP1</td>
<td>6</td>
</tr>
<tr>
<td>1.2</td>
<td>Domain organization of LTTRs using CbnR as the model</td>
<td>13</td>
</tr>
<tr>
<td>1.3</td>
<td>Ribbon representation of the tetrameric structure of CbnR</td>
<td>17</td>
</tr>
<tr>
<td>1.4</td>
<td>Ribbon representation of the DNA binding domain of CbnR</td>
<td>20</td>
</tr>
<tr>
<td>1.5</td>
<td>Diagram of a CysB monomer with a bound sulfate ion</td>
<td>26</td>
</tr>
<tr>
<td>1.6</td>
<td>Diagram of the DntR dimer with a bound acetate ion</td>
<td>29</td>
</tr>
<tr>
<td>2.1</td>
<td>BenM and CatM regulate benzoate degradation in ADP1</td>
<td>49</td>
</tr>
<tr>
<td>2.2</td>
<td>Expression of a chromosomal $\text{benA}::\text{lacZ}$ fusion in strains encoding CatM or variant regulators (CatM[V158M], encoded by catM5153, or CatM[R156H], encoded by catM3102).</td>
<td>63</td>
</tr>
<tr>
<td>2.3</td>
<td>Expression of a chromosomal $\text{catB}::\text{lacZ}$ fusion in strains encoding CatM or variant regulators (CatM[V158M], encoded by catM5153, or CatM[R156H], encoded by catM3102)</td>
<td>66</td>
</tr>
<tr>
<td>2.4</td>
<td>DNase I footprinting of BenM, CatM and CatM[V158M] at benA</td>
<td>73</td>
</tr>
<tr>
<td>2.5</td>
<td>Model of regulated benA expression</td>
<td>75</td>
</tr>
<tr>
<td>2.6</td>
<td>Comparison of sequences in the benA and catB promoter regions</td>
<td>83</td>
</tr>
<tr>
<td>3.1</td>
<td>Benzoate degradation in ADP1 is regulated by CatM and BenM and the effectors benzoate and muconate</td>
<td>98</td>
</tr>
<tr>
<td>3.2</td>
<td>Ribbon representation of BenM-EBD and CatM-EBD</td>
<td>111</td>
</tr>
</tbody>
</table>
Figure 3.3: Stereo views of the primary effector binding site and the secondary activation site

Figure 3.4: Multiple sequence alignment and secondary structure assignment

Figure 3.5: Schematic representation of the effector binding sites

Figure 3.6: Schematic model for the conformational changes in the full-length BenM or CatM tetramer in response to binding of its effector cis,cis-muconate

Figure 4.1: Benzoate degradation pathway in ADP1

Figure 4.2: Crystal morphologies of CatM- and BenM-EBD variants

Figure 4.3: Stereoview of part of the proposed tetrameric interface of BenM

Figure 4.4: Ribbon diagram of BenM(R156H/T157S) superimposed over BenM-EBD (PDB code: 2F6G) or BenM-EBD bound to muconate

Figure 4.5: Stereo representation of the residues adjacent to the location of the mutation in BenM-EBD and BenM(R156H/T157S)-EBD

Figure 4.6: Charge re-distribution of the molecular surface due to mutations in the BenM(R156H/T157S)-EBD structure

Figure 4.7: Comparison of structures of BenM, CatM and their variant EBDs

Figure 4.8: Stereo representation of an overlay of CatM-EBD, CatM-EBD bound to muconate and CatM(R156H)-EBD bound to muconate

Figure 4.9: Structural model of a compact form of a full-length CatM monomer using another LTTR, the CatM-EBD and CbnR structures as templates

Figure 5.1: Ribbon representation of the monomeric and dimeric structures of BenM-EBD structures A and B
Figure 5.2: Ribbon representation of BenM-EBD tetramer structure A in the P43_{22} space group (Panel A) and structure B tetramer in space group C22_{21} (panel B) ..........202

Figure 5.3: Interactions at the tetrameric interface of crystal structures A and B ......................207

Figure 5.4: Unit cell representations of BenM-EBD structures A and B ....................................212

Figure 5.5: Proposed schematic for the oligomerization of BenM in light of the CbnR structure (1IZ1) ..........................................................................................................................214
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Angstrom</td>
</tr>
<tr>
<td>α-CTD</td>
<td>C-terminal domain of the α-subunit</td>
</tr>
<tr>
<td>AU</td>
<td>Asymmetric unit</td>
</tr>
<tr>
<td>4HB</td>
<td>4-hydroxybenzoate</td>
</tr>
<tr>
<td>BD</td>
<td>Butanediol</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA binding domain</td>
</tr>
<tr>
<td>DNT</td>
<td>Dinitrotoluene</td>
</tr>
<tr>
<td>EBD</td>
<td>Effector binding domain</td>
</tr>
<tr>
<td>FOM</td>
<td>Figure of merit</td>
</tr>
<tr>
<td>HTH</td>
<td>Helix-turn-helix</td>
</tr>
<tr>
<td>HTS</td>
<td>High-throughput crystallization screens</td>
</tr>
<tr>
<td>LTTR</td>
<td>LysR-type transcriptional regulator</td>
</tr>
<tr>
<td>MAD</td>
<td>Multiple-wavelength anomalous diffraction</td>
</tr>
<tr>
<td>NAS</td>
<td>N-acetyl-l-serine</td>
</tr>
<tr>
<td>OAS</td>
<td>O-acetyl-l-serine</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>SAD</td>
<td>Single-wavelength anomalous diffraction</td>
</tr>
<tr>
<td>SBC-CAT</td>
<td>Structural Biology Consortium Collaboratory Access Team</td>
</tr>
<tr>
<td>SER-CAT</td>
<td>SouthEast Regional Collaboratory Access Team</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxylic acid</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction and Literature Review

Purpose of study

The expression of biodegradation pathways in microorganisms is tightly regulated at the transcriptional level. A thorough understanding of the regulatory mechanisms controlling biodegradation is crucial to the success of bioremediation approaches (28, 80). This dissertation focuses on the regulatory network involving two transcriptional regulatory proteins BenM and CatM, which control at least 15 ben and cat genes needed for benzoate degradation in Acinetobacter baylyi strain ADP1. CatM and BenM are in the large and diverse prokaryotic family of LysR-type transcriptional regulators (LTTRs) (73). Both regulatory proteins are 59% identical and play overlapping roles in the control of genes needed for benzoate degradation (17). Both proteins respond to a metabolite of benzoate degradation, cis,cis-muconate (hereafter referred to as muconate). BenM, unlike CatM also responds to benzoate. In addition, BenM responds to both muconate and benzoate in a synergistic fashion to activate gene expression (12). Recent structural studies, described here, have improved our understanding of the molecular basis of transcriptional control by these regulatory proteins in response to effectors and in relation to past mutational studies of LTTRs.

After a review in this chapter of the previous literature pertinent to this work, chapter 2 addresses the transcriptional regulatory differences and evolutionary relationship of BenM and CatM. The studies focus on the benABCDE operon in which both regulators exert significantly different effects. In the absence of BenM, CatM is not sufficient to
support growth on benzoate as the sole carbon source (21). These studies were conducted in light of three point mutations that independently increased CatM-mediated activation of the \textit{benABDCE} operon, thus enabling stains lacking BenM to grow on benzoate. These mutations generated variants with one amino acid change in CatM, [CatM(R156H) and CatM(V158M)]. The effects of both variants and BenM on the \textit{benA} promoter were determined in light of a third point mutation at the \textit{benA} promoter region that also increases CatM activation of the \textit{benABCDE} operon. These investigations highlight the importance of balanced expression from multiple promoters within the pathway.

The first structures of LTTRs bound to their cognate effectors are presented in chapter 3. Structures of truncated versions of CatM and BenM were solved with and without their effectors, muconate and benzoate. Truncated versions of these proteins represent about 70\% of the total protein and are known to be responsible for effector binding (18). As described in chapter 3, these domains are referred to as effector binding domains (EBD). Both effectors, muconate and benzoate bound in an inter-domain cleft in BenM-EBD. Benzoate, but not muconate, bound to an additional hydrophobic core within BenM-EBD. Comparative studies are interesting because these proteins share common functional traits but also have significant differences. These studies highlight important features of how both proteins respond to effectors to initiate transcriptional activation. These investigations also postulate a molecular basis for the synergism exhibited by BenM in response to both benzoate and muconate.

Continuing on the variant regulatory proteins presented in chapter 2, chapter 4 explains the structural basis for the observed altered regulation in these variants. In this chapter, the crystal structures of CatM(V158M) and CatM(R156H)-EBDs were solved at
1.80 Å and 2.45 Å respectively. In addition two independent crystal structures of a variant of BenM, BenM(R156H/T157S)-EBD were determined at 1.80 Å and 1.85 Å. This BenM variant allows the expression of the *ben* genes in the absence of effectors. In-depth analyses of these variant structures show how these mutations affect gene regulation by inducing conformational changes and affecting tetrameric interactions within the protein.

In light of the structural studies of chapters 3 and 4, two interesting unanswered questions are why LTTRs in general have insolubility problems and secondly, how their subunits interact to form tetramers and probably high-order oligomers. Chapter 5 of this dissertation answers these questions through the elucidation of two independent crystal structures of BenM-EBD. The crystal structures were obtained from different space groups, whose crystals were derived from high pH conditions. These structures provide a description of what is likely the tetrameric interface in this LTTR. The residues in this interface and their interactions that lead to the formation of a tetrameric molecule were identified. The studies show how certain conditions will favor oligomerization of the protein, thus leading to solubility problems.

*Acinetobacter baylyi* strain ADP1: History and general characteristics

Bacteria of the genus *Acinetobacter* are ubiquitous in nature and can be recovered from water, soil and living organisms. Bacteria of this genus are classified in the gamma subdivision of the proteobacteria. They are Gram negative, oxidase-negative, strictly aerobic and form non-motile cocci. Due to their nutritional versatility, they are able to use various carbon sources for growth and can be cultured in relatively simple media. In
1968, Baumann classified this group of organisms now known as *Acinetobacter* in the genus *Moraxella* (6). The basis for the classification included criteria such as nutritional properties, cell shape, G+C content of the genome and absence of flagella. Specific procedures were designed for the cultivation of strains of this genus from soil and water (6, 7). In recent times there has been a burgeoning interest in species of *Acinetobacter* due to their importance in many environmental and biotechnological applications, especially the degradation of toxic pollutants (1, 50, 75) and the cellular production of economically valuable products (32, 47). Other species have been studied because they are major causes of hospital-acquired infections worldwide and a public health problem in many countries (29, 53). A comparative sequence analysis of PCR-amplified 16S DNAs from 21 strains of *Acinetobacter* have shown that this genus is highly heterogeneous (40).

Within this genus, is the unique strain *Acinetobacter baylyi* ADP1 (formerly called *Acinetobacter* sp. or *Acinetobacter calcoaceticus* strain ADP1) (84). The strain was first isolated in 1969 by Elliot Juni (46). It was a mutant from a strain, designated BD4 (BD stands for butanediol) which was isolated from mineral medium with meso 2,3-butanediol as the sole carbon source. The BD4 strain produced abundant amounts of polysaccharide that hindered manipulation. A mutant from BD4, called BD413 was obtained by ultraviolet irradiation and later called ADP1 (46). *A. baylyi* ADP1 is a soil bacterium with a similar range of substrates that can be used as sole carbon and energy sources as two other well-studied bacteria, *Pseudomonas aeruginosa* and *Pseudomonas putida*. The genome of *A. baylyi* ADP1 has recently been sequenced and has a single, circular chromosome of 3.7 Mb with an average G+C content of 40.3% that encodes 3325
predicted proteins (3). *A. baylyi* is highly competent for natural transformation which extends to both plasmid DNA and linear fragments (46, 60). There is also a strong tendency towards homology-directed recombination (26). As reviewed in a recent article (89), these properties have been instrumental in developing *A. baylyi* strain ADP1 as a model organism for biochemical, genetic and physiological studies. The flexibility and versatility of ADP1 makes it an ideal organism for genome engineering and automation of complex strain construction (56). The natural competence of ADP1 has also allowed the development of this organism as a powerful model system for the study of chromosomal rearrangements and DNA amplification engineering (69, 70). The recent genome analysis of *A. baylyi* strain ADP1 showed that 20% of the genes could be associated with catabolic functions with all of these genes located in five major "islands of catabolic diversity" (65).

**Aromatic compound degradation by *Acinetobacter***

A striking feature of this genus is its capability to grow on a wide array of aromatic compounds. The metabolic pathway for the degradation of aromatic compounds is the β-ketoadipate or ortho-cleavage pathway (Fig. 1.1A). This pathway has been identified primarily in soil and aquatic/marine microorganisms. Enzyme studies and amino acid sequence data show that the pathway is highly conserved in diverse bacteria and fungi. This pathway has been well characterized in *A. baylyi* strain ADP1 with respect to the enzymatic steps and encoding genes as summarized in two reviews (35, 89). The first phase of this pathway is the conversion of structurally diverse compounds by chromosomally encoded enzymes into one of two dihydroxylated intermediates, catechol
FIG. 1.1. The β-ketoadipate pathway in strain ADP1 (Panel A). Compounds and transcriptional regulators relevant to this study are shown in bold. Brackets indicate genes whose expression can be controlled by the transcriptional regulators shown. PobR and PcaU are IclR (GylR)-type regulators. BenM and CatM are LysR-type regulators. Compounds which are circled are important effectors of transcription control by BenM and CatM. Panel B (not drawn to scale) shows the arrangement of the \textit{ben} and \textit{cat} operons involved in benzoate degradation arranged in 20 kbp chromosomal cluster. The relative positions and transcriptional directions are noted. The diamonds represent transcription initiation sites for BenM and/or CatM. BenM is the major regulator of the \textit{benABCDE} operon although CatM can also activate transcription from the \textit{benA} promoter. CatM is not sufficient to allow a strain lacking BenM to grow on benzoate at wild-type growth rates. CatM is the major regulator of the \textit{catBCIJFD} operon, although BenM can regulate \textit{cat}-gene expression under some conditions.
A. Protocatechuate Branch

1. 4-hydroxybenzoate (4HB)
2. Protocatechuate

- PobR
  - PobA
  - pcaGH
  - pcaB
  - pcaC
  - pcaD
  - pcaIJ
  - pcaF

B. Catechol Branch

1. Benzoate
2. Catechol
3. cis,cis-muconate

- antABC
- CatM

B. BenPK benM benABCDE catA ORF 1,2 catM catBCIJFD

- BenM
- BenM
- BenM
- CatM
or protocatechuate (Fig. 1.1A). The second phase of degradation includes ring fission by substrate-specific dioxygenases with subsequent reactions leading to the generation of tricarboxylic acid cycle intermediates. Catechol and protocatechuate are cleaved by catechol 1,2-dioxygenase and protocatechuate 3,4-dioxygenase respectively. The cleavage products, muconate and 3-carboxy-muconate are transformed via the common intermediate, \( \beta \)-ketoadipate enol-lactone to the tricarboxylic acid cycle intermediates succinyl-CoA and acetyl-CoA. This dissertation primarily focuses on the catechol branch of the pathway in which compounds such as anthranilate, benzoate, benzyl acetate, benzoate, salicylate are channeled through catechol (11, 35, 42, 43).

**Regulation of genes needed for benzoate degradation**

*Arthrobacter baylyi* strain ADP1 has a BenM and CatM dual regulatory system (Fig. 1.1B). This regulatory system serves as a good model for identifying key interactions among proteins, inducers and DNA. Both regulatory proteins play equal roles in governing the regulation of the *benPK* promoter (17, 22). However, BenM plays the primary role in governing the activation of the *benABCDE* promoter (21). Although CatM can activate the *benABCDE* promoter, it does not activate it to high enough levels to allow a strain lacking BenM to grow on benzoate as addressed in chapter 2. In the absence of CatM, BenM activates the expression of the *cat* genes but allows only very slow growth on benzoate (31). CatM and BenM regulate expression of the *catA* gene, required for the conversion of catechol to muconate (72). CatM regulates the expression of the *catBCLJFD* genes needed for the subsequent degradation of muconate to tricarboxylic acid cycle intermediates (58). In view of the toxicity of muconate to cells, it is explicable
that the genes needed for the formation and degradation of this toxic metabolite would be tightly regulated and transcribed from separate promoters. CatM was shown to bind to sequences upstream of catA and catB (20, 72). *In vivo* footprinting studies have defined key regulatory sequences upstream of catB by showing that CatM interacts with an LTTR consensus binding site centered approximately 65 nucleotides upstream of the catB transcriptional start site (72).

A single round *in vitro* transcription assay was developed for benA (12). This assay showed that BenM activates benA transcription in response to benzoate and muconate. Both metabolites had a synergistic effect on BenM activated transcription. The comparative structural analyses of BenM and CatM described in this dissertation provide a model for key structural features required for transcriptional synergism in BenM but not CatM.

**Transcriptional repression by BenM and CatM**

BenM and CatM like many LTTRs can repress their own expression. Both regulatory proteins have been shown to repress benA expression in the absence of effectors (12). DNase I footprinting studies have shown that BenM or CatM can protect an extended region of the benA operator/promoter region from DNase I cleavage (20). The protected region contains three LTTR consensus sites as described in chapter 2. Each site should represent a binding region for a BenM or CatM dimer. One of these sites (site 3) directly overlaps with the -10 promoter region. In the absence of effectors, site 3 is protected and apparently prevents RNA polymerase from binding to initiate transcriptional activation.
BenM and CatM have also been shown to repress the genes needed for 4-hydroxybenzoate (4HB) metabolism during growth on benzoate (10). 4HB is degraded through the protocatechuate branch of the β-ketoadipate pathway (Fig. 1.1A). The pobA-encoded enzyme catalyzes the first step in the degradation of 4HB, a compound consumed rapidly as a sole carbon source. In ADP1, benzoate is consumed in preference to 4HB (31). It was shown that BenM and CatM could bind within pcaU, a gene involved in activating the pca genes (pcaIJFBDKCHG) needed for the degradation of protocatechuate (10). This preference is regulated by BenM and CatM, which may delay 4HB catabolism by repressing the expression of pcaK, a gene which encodes a transport protein that affects uptake of 4HB (24). PcaU activates expression of pcaK, thus the repression of pcaU would result in the absence of pcaK expression. One of my preliminary projects in the Neidle lab was to examine the role of pcaK deletion on pobA expression using a pobA::lacZ transcriptional fusion. The results supported the regulatory model in which BenM and CatM mediate the preference for benzoate over 4HB by regulating the transport of 4HB.

**Muconate levels in the cell affect CatM-mediated regulation**

Strains in which benM has been disrupted do not grow on benzoate as sole carbon source denoting that a functional copy of catM is not adequate for expression of the benABCDE operon to levels high enough to sustain growth on benzoate. Mutants that acquire the ability to grow on benzoate as the sole carbon source arise spontaneously at a frequency of approximately $10^{-5}$ (20). Two such mutants, one with a mutation in catM and the other with a mutation in the benA promoter are characterized in chapter 2. Four
other mutations were in the *catB* gene and caused sufficient *benA* expression to enable growth on benzoate (23). The *catB* gene encodes muconate cycloisomerase that catalyses the conversion of muconate to muconolactone (Fig.1.1A). Three of the four CatB variants were purified and shown to have reduced CatB activity (23). Another mutation was located in the *catB* operator/promoter region in which there was an extra T nucleotide within the LTTR consensus binding site upstream of the *catB* transcriptional start site.

One of my first projects on entering the Neidle lab was to determine the effect of this mutation on *catB* expression. The studies showed that the mutation reduced the level of muconate-inducible *catB* expression by about 7-fold relative to that of the wild-type promoter. The results suggest that the operator/promoter mutation causes a reduction in the expression of the *catB* gene leading to lower levels of muconate cycloisomerase, the enzyme which degrades muconate. This reduced activity should increase the intracellular concentrations of muconate. The increased muconate levels would increase CatM-mediated expression of the *ben* genes, thus allowing a strain lacking *benM* to grow on benzoate. Collectively, these results support the important role of metabolite concentrations in controlling benzoate degradation in ADP1 by LTTRs.

**Structure and function of LTTRs**

Numerous bacterial transcription regulatory proteins bind DNA via a helix-turn-helix (HTH) motif (39). These proteins are very diverse, but for convenience, they can be categorized into different families on the basis of sequence similarity. About twelve families of regulatory DNA-binding proteins have been identified (67). One such family is the LysR family of transcriptional regulators. This family may be the most common
type of positive transcriptional regulators in prokaryotes (36, 61, 73). LTTRs, first
described by Henikoff et. al. are present in diverse bacterial genera, archaea, and algal
chloroplasts. Biochemical studies and sequence analyses of various LTTRs show that
they are similarly sized molecules (~300–400 amino acids). LTTRs control genes not
only involved in aromatic compound catabolism but also in a variety of other
physiological processes such as amino acid synthesis, oxidative stress, carbon fixation,
nitrogen fixation and bacterial virulence. Consequently, a thorough understanding of
LTTRs has broad significance.

Most LTTRs, while activating expression of target genes, repress their own
expression, frequently by the use of divergent promoters as observed for BenM (12).
DNA sequences recognized by LTTRs contain a consensus binding site with “T-eleven
nucleotides-A” in which the “T” and “A” reside in short segments with dyad symmetry.
LTTRs generally act as tetramers in their active form (36).

Members of the LysR family have conserved and similarly organized functional
domains (Fig. 1.2). Extensive mutational analyses have been done on many members of
the family such as OxyR (48, 49), CysB (51, 52), CbbR (25) and NahR (62, 74). The
results have pinpointed regions of the proteins that are important in DNA binding and
inducer response/recognition (73). However, the recent structures of two LTTRs, CbnR
(57) and DntR (76) have been crucial in improving our understanding of domain
organization and the oligomeric state of LTTRs with respect to transcriptional regulation.
CbnR regulates the cbnABCD operon for the degradation of 3-chlorobenzoate from
plasmid pENH91 of Ralstonia eutropha NH9 (59). DntR regulates the genes needed for
the degradation of 2,4-dintrotoluene in Burkholderia sp. strain DNT (76). The three
**FIG. 1.2.** Domain organization of LTTRs using CbnR as the model. Domains are colored according to function. The red areas correspond to the DNA binding domain which contains the HTH motif (colored black). The area colored yellow corresponds to the linker region connecting both the DNA binding domain and regulatory domain. The area colored blue corresponds to the regulatory domain or effector binding domain.
(1-57) HTH
DNA Binding Domain

(59-87) linker helix

(90 - C-terminal residue) Effector binding domain/Regulatory domain
dimensional structures of the first and only well characterized full-length LTTR is CbnR (57). It is also currently the only full-length LTTR for which its structure is known. However, a modeled structure of the full-length DntR has been reported (76). Crystals of the full-length CysB (86) and cofactor binding domain of Cbl (a CysB-like protein) (77) have also been reported but their structures are yet to be resolved. The regulatory domain of LTTRs based on the structure of CbnR includes residues 90 to the C-terminal end of the protein. This region (the regulatory domain or effector binding domain) shows great variability among family members compared to the DNA binding domain. Structural analyses of the regulatory domains of OxyR (14), CysB (82), DntR (76) and CbnR (57) reveal that this domain can be further divided into two sub-domains usually designated domains I and II. As described more fully in this dissertation, domain I of BenM is composed of amino acid sequences corresponding to residues 90 - 161 and 268 - 304, whereas, domain II is composed of one contiguous sequence corresponding to residues 162-267. Mutational analyses resulting in residue changes in the regulatory region have been shown to affect mainly inducer response among members of this family (51, 73). Mutations in the regulatory region specifically elicited constitutive activity of regulatory proteins or affected its response to inducers. The predominant regions involved in these different effects seem to fall between residues 90 to 274 in AlsR (71), AmpR (4, 5), CatM (58) CbbR (25), CysB (51), NahR (74), NodD (13), OccR (2), OxyR (49) and XapR (44).

Although mutational analyses suggest that residues 90 to the C-terminal residue are important in effector recognition/response, these studies fail to explain the mechanism by which residues in this region might affect inducer response and recognition. For
instance, are these residues involved in direct or indirect binding with the effector? Do these mutations mimic effector binding? Or do these residues simply affect the way these proteins dimerize or tetramerize? It is difficult to draw specific conclusions from these studies because of the paucity of structural data. To date, no crystal structures of LTTRs have been solved with their cognate effectors. This dissertation describes the first crystal structures of LTTRs bound to their cognate effectors. Another challenge in interpreting some of the data, is that residues from different proteins that are in the same position in the amino acid sequence differ in position structurally on a three-dimensional level. The crystal structures presented in this dissertation provide a foundation for evaluating LTTR-effector interactions.

The structure of CbnR

CbnR is in the sub-family of well characterized LTTRs to which BenM and CatM belong. It regulates the cbnABCD operon for the degradation of 3-chlororobenzoate from plasmid pENH91 of Ralstonia eutropha NH9 (59). It might also respond to muconate and halogenated-muconate inducer compounds, although precise studies of effector response have not yet been published (59). Determination of the CbnR structure was done using a molecular replacement method and multiple anomalous dispersion diffraction phasing at a resolution of 2.2 Å and R-factor of 21.7%. The CbnR homo-tetramer, which is ellipsoidal in shape, can be divided into two domains, the DNA binding domain (residues 1-58) and the regulatory domain (residues 90-294). Both domains are connected by a linker helix (residues 59-87), made up of 29 amino acid residues (Fig. 1.3).
**FIG. 1.3.** Ribbon representation of the tetrameric structure of CbnR (57) (A). Subunits 1, 2, 3 and 4 are shown in green, pink, yellow and red respectively. The locations of the DNA-binding domains and coiled-coil linkers are shown. The 2-fold axis of symmetry in the molecule is shown as a black arrow on top. Panels B and C show that the subunits fold into an extended form (shown in B) or a compact form (shown in C). Regulatory domain I, II, linker helix and the DNA binding domain are colored yellow, grey, red and blue respectively. The helices are shown as cylinders. The figures were constructed using pymol (27).
The four identical subunits (monomers) that make up the tetrameric CbnR can adopt two distinct conformations in the tetramer. Two subunits adopt an extended form with a corresponding angle of 130° between the linker helix and the regulatory domain (Fig. 1.3B). The other two subunits adopt a compact form in which the angle between the linker helix and the regulatory domain is 50° (Fig. 1.3C).

The regulatory domain of CbnR can be divided into two sub-domains, domains I and II. Domain I is made of two separate regions of the polypeptide chain, residues 90 - 161 and residues 265-294 comprised of five β-strands, three α-helices, and one 310–helix. Domain II is made up of residues 164-259 and is composed of five β-strands, three α-helices, and two 310–helices. Although domain II has an extra 310-helix, the folds of both domains are similar to one another.

The DNA binding domain of CbnR was determined to be a winged helix motif with three α-helices and two β-strands and extended loop structures called wings (57). This motif consists of two helices packed at angles of ~ 120° joined by a tight four-residue turn (9). Each monomer of a winged helix protein contains a helix-turn-helix motif followed by one or two β-hairpin wings. The loop structures ("wings") are designated wing1 and wing2 as shown in fig. 1.4. These wings provide considerable structural flexibility in DNA binding and recognition (39). The winged helix motif is a variation of the helix-turn-helix motif (8). The closest related structure to the DNA binding domain of CbnR was that of the ModE protein from *Escherichia coli*, involved in molybdate-dependent transcriptional regulation (34).

Mutational studies suggest that the residues within the DNA binding domain are necessary for DNA binding and recognition (44, 48, 51, 52). The DNA binding domain
FIG. 1.4. Ribbon representation of the DNA binding domain of CbnR (57). The helices of the HTH motif are $\alpha_2$ and $\alpha_3$ respectively. Wing1 and Wing2 are loops that form the extended $\beta$-sheath structures of the winged helix motif. The residues at the turn region of the HTH are shown in stick representation. Residues at this turn region of CysB have been shown to contact the C-terminal domain of the $\alpha$-subunit of RNA polymerase (52).
residues at the turn region of the HTH
has also been implicated in the interaction of LTTRs with RNA polymerase during transcription. Most bacterial transcriptional regulators other than LTTRs contact the $\alpha$-subunit and the $\sigma$ subunits of RNA polymerase to initiate transcription (37). Interactions of LTTRs with RNA polymerase have been studied in protein such as CysB (52), OxyR (79), CatR (54), GcvA (45), NahR (63) and MetR (30). The studies showed that these proteins mainly interacted with the C-terminal domain of the $\alpha$-subunit ($\alpha$-CTD) of RNA polymerase. The CysB investigations by Lochowska et al., revealed the precise residues of CysB involved in the interaction with RNA polymerase. Through the process of targeted alanine scanning of the DNA binding domain of CysB, residues lying in the turn region of the helix-turn-helix motif were shown to be at least one activating region of CysB. This activating region consisted of surface residues and was shown to make contact with RNA polymerase (the corresponding residues in CbnR are shown in fig.1.4).

In another set of experiments, the authors used a LexA-based two hybrid system designed to determine what subunit or region of RNA polymerase makes physical contact with CysB. The results showed that CysB interacted with the $\alpha$-CTD of RNA polymerase. Furthermore, by the use of variants of the $\alpha$-CTD of RNA polymerase, the authors also pinpointed what region of $\alpha$-CTD makes contact with the activating region of CysB. This region identified is known as the "273 determinant" (41) and comprises residues K271, E273 and neighboring residues.

The mechanism by which inducer binding appears to change the conformation of the regulatory domain, and subsequent change in the global conformation of the tetramer is still unclear because none of the CbnR crystal structures were derived with the effectors present. Although the presumed effector molecule of CbnR, muconate was added to the
crystallization solution, no densities of the effectors were seen in the structures. This could be attributed to the very high concentration of NaCl used in the crystallization attempts (4.3 M), which may compete with inducer binding to the protein. Secondly, the concentration of muconate added to the crystallization trials was relatively low (between 3 - 6 mM), compared to the concentration used for the CatM-EBD and BenM-EBD structures presented in this dissertation (≥ 100 mM). One of the goals of this dissertation is to describe how the LTTR family of regulators responds to their cognate effectors to activate transcription of target genes.

**LTTRS involved in aromatic compound degradation**

As summarized in a review by Diaz and Prieto (28), LTTRs associated with aromatic compound degradation are diverse. Phylogenetic analysis show that both BenM and CatM belong to a subfamily of well-characterized LTTRs, in which many members respond to muconate and halogenated-muconate. Examples of regulatory proteins in this subclass that degrade aromatic compounds include: TfdR, controlling chlorocatechol metabolism from plasmid pJP4 of *Ralstonia eutropha* (85); TcbR controlling degradation of 1,2,4-trichlorobenzene, encoded by plasmid pP51 of *Pseudomonas* sp. strain P51 (83) and CatR controlling catechol metabolism of *Pseudomonas putida*. (15). ClcR regulates the *clcABD* operon (19); TcbR regulates the *tcbCDEF* operon in plasmid pP51 of *Pseudomonas* sp. strain P51 (83); CatR regulates the *catBCA* operon of the catechol branch in response to muconate in *P. putida* (16) and CbnR regulates the *cbnABCD* operon for the degradation of 3-chlorobenzoate from plasmid pENH91 of *Ralstonia eutropha* NH9 (59).
Interactions of LTTRS with ligands

Most LTTRs identified to date generally require low molecular weight ligands to act as co-activators in the positive control of transcription at target promoters (54, 73). All LTTRs involved in aromatic compound degradation involve such effectors (81). The structures of the effector binding domains of BenM and CatM presented in this dissertation are very similar to bacterial periplasmic receptors. These receptor proteins consist of two distinct globular domains bisected by a cleft or groove in which substrates or effectors can bind within the cavity region causing the domain to contract (68).

The LTTR effectors are frequently a pathway substrate such as benzoate and muconate for BenM (12); muconate for CatM (72) and CatR (16); 2-chloro-muconate for ClcR (19) and probably CbnR (59); chloro-muconate for TFdR/S (85); β-Carboxymuconate and γ-carboxymuconate for PcaQ (64); salicylate for DntR (76) and NahR (74). Most LTTRs also bind to their target promoter DNA under both inducing and non-inducing conditions. Interactions with cognate effectors usually cause change in the DNA/protein interactions as seen in the DNase I footprints of many LTTRs. Examples of LTTRs in which DNase I footprinting studies have shown changes in protein/DNA interactions in response to effectors include CysB (52), BenM (12), CatR and ClcR (55). These changes seem to involve the translocation of the tetrameric protein from one binding site to another as described more completely in chapter 2 of this dissertation concerning the activation of the benABCDE operon promoter by CatM.

Although the crystal structures of CysB, DntR, OxyR and CbnR have been elucidated, none of these structures were solved with their cognate effector bound to the protein. The crystal structure of the chymotryptic fragment of Klebsiella aerogens, CysB
was solved by multiple isomorphous replacement and multi-crystal averaging to 1.8 Å resolution (87). CysB controls numerous genes involved in bacterial sulfur assimilation via cysteine biosynthesis. Positive control of CysB requires the effectors O-acetyl-L-serine (OAS) or N-acetyl-L-serine (NAS) (38, 51). Although crystals of CysB with NAS have been reported, the structures of these complexes are yet to be solved (86). The structure of the chymotryptic fragment of CysB contained a sulfate molecule. The sulfate molecule was presumed to have come from the fractionation step during the protein purification process in which ammonium sulfate is used as the precipitant. The sulfate binding site was found within the inter-domain cleft of a monomer (fig. 1.5). As shown in figure 1.5B, the sulfate anion interacts extensively with the protein main chain, side chain groups and surrounding water molecules. Protein residues involved in these interactions were T100, T102, T149, T202, Q103 and two water molecules. Other residues that line the cavity but are not involved in any interactions with the sulfate ion included H101, G128, G129, P131, E150, Y197 and T225. The implications of this finding by Tyrell et al and the structural characterization of OxyR three years later provided a hypothesis that this sulfate binding region could be a potential target for effectors of LTTRs in general (14). Recently, the sites for effector binding were further investigated in the structural characterization of another LTTR, DntR (76).

**Structural characterization of the inducer binding cavity of DntR**

DntR regulates the genes involved in oxidative degradation of 2,4-dinitrotoluene (DNT) by *Burkholderia* sp strain DNT (33, 78). This strain can use DNT as the sole carbon source (78). Although there are no structures of DntR bound to the cognate
FIG. 1.5. Diagram of a CysB monomer with a bound sulfate ion (87). Panel A is a ribbon representation of the CysB monomer with respect to the two major domains of the effector binding domain. Residues 88-164 and 268-324 comprise domain I colored green. Domain II is colored pink and is comprised of residues 165 - 267. The sulfate molecule is shown in red, located in the inter-domain cleft between both domains. Panel B shows the interaction of sulfate with residues in both domains. The residues that line the effector binding cavity are shown in ball-and-stick representation with carbon atoms colored pink for domain II and green for domain I. Sulfate is also shown in ball and stick with oxygen atoms colored red and carbon atoms colored orange. The backbone and ribbons are colored according to both domains as in Panel A. Direct and water-mediated hydrogen bonding interactions are shown as broken black lines. Water molecules are shown as small spheres. The figures were constructed using pymol (27).
effectors (DNT or salicylate), the structure of DntR bound with acetate or thiocyanate was recently solved to resolutions of 2.3 Å and 2.6 Å respectively (76). The location of these compounds in the DntR structure was in the inter-domain cleft of two domains. This was very similar to the location of the sulfate residue in the structure of CysB (82).

As reported by Smirnova et al., the cavity was found at the interfaces of sub-domains I and II (Fig. 1.6A) (76). The entrance to the cavity of DntR was polar (T104, L151, H169 and H206) while the surface of the interior was mostly hydrophobic (L106, G107, Y110, F111, G152, F167, R248 and I273). In the acetate bound DntR structure (Fig. 1.6A), one of the oxygen atoms interacted with T104 and two other water molecules, one of which was interacting with the main-chain amide group of T104 and one with the carbonyl oxygen of L151. The other oxygen atom of acetate interacted with N atoms of H169 and H206. For the thiocyanate bound DntR structures, the thiocyanate ion, interacted with the protein through similar ionic and hydrophobic interactions as seen in the acetate bound structures. The authors, using these locations of acetate and thiocyanate were able to model the cognate effectors salicylate and DNT in this effector-binding cavity. A superimposition of the structures of DntR and CysB on BenM or CatM show that these cavities are structurally similar in size and location. However they differ in the kind of residues surrounding the pocket. This could explain the specificity of LTTRs for certain molecules as effectors in activating transcription. Interestingly, the inter-cleft binding pockets for CatM and BenM are almost identical as presented in chapter 3.

Although the full-length DntR protein was crystallized, the electron density for the major part of the DNA binding domain is poor. However, the authors were able to position, but not refine, the linker helices and the two helices in the DNA binding domain
FIG. 1.6. Diagram of the DntR dimer with a bound acetate ion (76). Panel A is a ribbon representation of the DntR dimer. Domains I and II of the effector binding domain are shown in blue and yellow respectively. Domain I consists of residues 90 - 181 and 272 - 301 while domain II consists of residues 169 - 271. (B) The residues that line the effector binding cavity are shown in ball-and-stick representation. Oxygen atoms are colored red and nitrogen atoms colored blue. Acetate is also shown in ball and stick with oxygen atoms colored red and carbon atoms colored pink. Direct and water-mediated hydrogen bonding interactions are shown as broken black lines. Water molecules are shown as red spheres. The figures were constructed using pymol (27).
for both DntR monomers found in the asymmetric unit of the crystals. This homotetramer consists of a dimer of dimers in which the dimer–dimer interface is very similar to that of CbnR. The tetrameric interface is also very similar to those observed for the high pH BenM-EBD structures presented in chapter 5 in which the tetrameric interfaces of CbnR, DntR and BenM were comparatively analyzed in light of the insolubility problems associated with this family of proteins.

The structures of BenM and CatM

The tetrameric nature of most LTTRs, including BenM and CatM make them ideal for binding at two positions simultaneously on the promoter. The structure of CbnR as a dimer of dimers supports the model in which LTTRs can bind to a long stretch of DNA of approximately 50 to 60 bp. The interactions of LTTRs with target promoter regions have shown that a sharp bend in the DNA is induced upon binding. Circular permutation gel shift assays have shown that LTTRs such as CatR and OccR induce an approximately 60° bend in the promoter in the absence of effectors (66, 88). However, in the presence of effectors this bend is relaxed to about 45°. These changes are based on effector interactions with the LTTRs to induce changes in the promoter DNA that enable transcriptional activation. The changes that have been observed for BenM and CatM interactions with the benA promoter region are outlined in chapter 2.

The crystal structures of the effector binding domains of CatM and BenM provide a framework for the structural conformational changes of these regulatory proteins in response to their cognate inducers in activating transcription. These structures also highlight the basis for the synergism in response to benzoate and muconate by BenM as
presented in chapter 3. Collectively, this dissertation clarifies the roles of BenM and CatM in regulating the genes needed for benzoate degradation in ADP1.
REFERENCES


17. \textbf{Clark, T. J., C. Momany, and E. L. Neidle}. 2002. The \textit{benPK} operon, proposed to play a role in transport, is part of a regulon for benzoate catabolism in \textit{Acinetobacter} sp. strain ADP1. \textit{Microbiology}. \textbf{148}:1213-23.


transcription of the distal *ben* genes and contribute to a complex regulatory circuit in *Acinetobacter* sp. strain ADP1. *J. Bacteriol.* **182**:7044-52.


30. **Fritsch, P. S., M. L. Urbanowski, and G. V. Stauffer.** 2000. Role of the RNA polymerase alpha subunits in MetR-dependent activation of metE and metH:
important residues in the C-terminal domain and orientation requirements within RNA polymerase. *J. Bacteriol.* **182**:5539-50.


synthase/acyl-CoA:diacylglycerol acyltransferase mediates wax ester and 
triacylglycerol biosynthesis in *Acinetobacter calcoaceticus* ADP1. *J. Biol. Chem.* 
**278**:8075-82.

the redox-sensitive transcriptional regulator OxyR: regions important for DNA 

analysis of the redox-sensitive transcriptional regulator OxyR: regions important for 


important for DNA binding, inducer response, oligomerization, and positive control. 

Bolewska, and M. M. Hryniewicz. 2004. Identification of activating region (AR) of 
*Escherichia coli* LysR-type transcription factor CysB and CysB contact site on RNA 


family member, CysB: a familiar fold with a surprising subunit arrangement.

*Structure.* 5:1017-32.


CHAPTER 2

CATM REGULATION OF THE BENABCDE OPERON: FUNCTIONAL DIVERGENCE OF TWO LYSR-TYPE PARALOGS IN ACINETOBACTER BAYLYI ADP1*

† OCE and L.S.C.-H contributed equally to this report.
Reprinted here with permission of the publisher, 6/2/06.
ABSTRACT

Two LysR-type transcriptional regulators, BenM and CatM, control benzoate consumption by the soil bacterium *Acinetobacter baylyi* ADP1. These homologs play overlapping roles in the expression of multiple genes. This study focuses on the *benABCDE* operon that initiates benzoate catabolism. At this locus, BenM and CatM each activate transcription in response to the catabolite *cis,cis*-muconate. BenM, but not CatM, additionally responds to benzoate as an effector. Regulation by CatM alone is insufficient for growth on benzoate as the sole carbon source. However, three point mutations independently increased CatM-activated *benA* transcription and enabled growth on benzoate without BenM. Two mutations generate variants with one amino acid change in the 303-residue CatM, CatM[V158M] and CatM[R156H]. These substitutions affected *benA* regulation differently than that of *catB*, another CatM-regulated gene involved in benzoate catabolism. In relation to CatM, CatM[V158M] increased *cis,cis*-muconate dependent transcription of *benA* but decreased that of *catB*. CatM[R156H] increased effector-independent expression of *catB* compared to CatM. In contrast, *cis,cis*-muconate was required with CatM[R156H] to activate unusually high *benA* expression. Thus, induction by *cis,cis*-muconate depends on both the sequence of CatM and the promoter. A point mutation at position –40 of the *benA* promoter improved CatM-activated gene expression and altered regulation by CatM[R156H]. BenM and CatM bound to the same locations on *ben*-region DNA. The frequency with which spontaneous mutations allow CatM to substitute for BenM might predict that one regulator would be sufficient for controlling benzoate consumption. This prediction is discussed in light of current and previous studies of the BenM-CatM regulon.
INTRODUCTION

In the soil bacterium Acinetobacter baylyi (formerly sp.) strain ADP1, BenM and CatM are homologous transcriptional regulators involved in aromatic compound degradation (Fig. 2.1) (8, 12, 29). These proteins are 59% identical in sequence, and both respond to a metabolite formed during benzoate consumption, cis,cis-muconate (hereafter referred to as muconate; Fig. 2.1A). However, unlike CatM, BenM also responds to benzoate as an effector. These two regulators jointly activate more than a dozen chromosomal ben and cat genes involved in benzoate catabolism (Fig. 2.1B). Furthermore, during growth with benzoate as a carbon and energy source, BenM and CatM repress genes needed to degrade alternative aromatic compounds (4).

In mutants lacking catM or benM, there is little effect on the expression of some genes such as catA or benPK (8, 12). At these loci, one regulator compensates for the loss of the other. This redundancy raises questions about the need for both regulators. To understand the evolution and retention of the benM and catM paralogs, these studies focused on the benABCDE operon. At this locus the regulators exert markedly different effects. BenM represses benA transcription in the absence of its effectors (6, 12). In response to muconate or benzoate, BenM activates ben-operon transcription. Furthermore, both effectors together cause a synergistic increase in BenM-activated transcription (6).

CatM activates benABCDE transcription in response to muconate but not benzoate (11, 12). As assessed with a benA::lacZ fusion in a benM mutant, CatM with muconate activates expression at a level 7-fold below that of BenM with muconate and 21-fold below that of BenM with both effectors (11, 12). Regulation by CatM is
FIG. 2.1. BenM and CatM regulate benzoate degradation in ADP1. Catabolism depends on enzymes encoded by the *ben* and *cat* genes (A). These chromosomal genes are in an approximately 20 kbp cluster (B, not drawn to scale). BenM and CatM regulate transcription initiation in four regions (diamonds 1-4). The functions of two open reading frames downstream of *catA* (orf1 and 2) are unknown, but they are not expressed during growth on benzoate. In the intergenic *benMA* region (C), there are three potential binding sites for BenM and CatM, as described in the text. Site 1 exactly matches the consensus sequence (underlined) of LysR-type regulators within a subclass to which BenM and CatM belong (10, 31). The sequences in Sites 2 and 3 differ from the consensus by a single nucleotide that reduces the extent of dyad symmetry. Above the DNA sequence, the transcription initiation site (+1) and promoter regions (-10, -35) are shown for *benA*. The initiation site for the divergently transcribed *benM* is also indicated (+1 below an arrow). A point mutation (circled A) increases the ability of CatM to activate transcription of the *benABCDE* operon (12).
Increased muconate-inducible activation of benABCDE by CatM

A

Benzoate → Catechol → cis,cis-Muconate → Succinyl-CoA + Acetyl-CoA

B

1. benPK 2. benABCDE 3. catA 4. catM

C

Increased muconate-inducible activation of benABCDE by CatM
insufficient to support growth on benzoate in mutants lacking BenM. However, CatM alone activates high-level transcription from the other promoters involved in benzoate consumption (\textit{benP, catA, catB}, Fig. 2.1). Therefore, mutations that increase \textit{ben}-operon expression permit benzoate to be consumed without BenM. For example, a point mutation in the –10 region of the \textit{benA} promoter enables a \textit{benM} mutant to grow on benzoate (12). Without inducers, this mutation increases \textit{benA} expression to a level 4-fold higher than that induced from the wild-type promoter by CatM with muconate (12).

Here, mutations that increase CatM-activated transcription from the \textit{benA} promoter were studied to identify constraints that normally prevent CatM from serving as the sole regulator of benzoate catabolism in ADP1. In the absence of CatM, BenM permits benzoate to be consumed very slowly (29). However, regulation by BenM in a \textit{catM} mutant causes abnormally high levels of muconate to accumulate (12, 15). This observation suggests limitations in the ability of BenM to activate transcription from the \textit{catB} promoter. BenM and CatM are members of the LysR-type family, a large group of homologous regulators that control diverse functions (31). The ADP1 regulators belong to a subclass of this family involved in the catabolism of aromatic compounds and pollutants in numerous bacterial genera (16, 32). Thus, the BenM-CatM regulon may serve as a good model for understanding complex regulatory circuits involved in biodegradation.

This report describes a variant, CatM[V158M], that enables growth on benzoate without BenM. CatM[V158M] was compared to CatM[R156H], a variant that activates \textit{cat}-operon expression without muconate (23). Additionally, the interactions of CatM with the \textit{benA} promoter region were compared to those of BenM at the same region. The
importance of the DNA sequence was explored in further investigations of a point mutation in the \textit{benA} promoter that increases CatM activation of \textit{benA}. These experiments help elucidate the threshold level of \textit{benABCDE} operon expression needed for growth on benzoate. Furthermore, they highlight the importance of balanced expression from multiple promoters within the pathway. The accumulation of toxic metabolites may be a key factor in the evolution of this regulatory scheme.

**MATERIALS AND METHODS**

**Bacterial strains and growth conditions**

\textit{Acinetobacter} strains (Table 2.1) were derived from ADP1, recently re-classified as an \textit{A. baylyi} strain (33, 37). Bacteria were cultured in Luria-Bertani broth or minimal medium at 37°C (12). \textit{Escherichia coli} DH5α (Invitrogen) was used as a plasmid host. \textit{E. coli} BL21(DE3) (Stratagene) was used to express and purify BenM and CatM. Carbon sources were added at the following final concentrations: 3 mM benzoate, 3 mM muconate or 10 mM succinate. Antibiotics were added as needed at the following final concentrations: ampicillin, 150 µg/ml for \textit{A. baylyi} or 50 µg/ml for \textit{E. coli}; kanamycin, 25 µg/ml; streptomycin 13 µg/ml; spectinomycin 13 µg/ml; and tetracycline 13 µg/ml. For growth curves, succinate-grown colonies were used to inoculate 5-ml cultures for overnight growth with succinate as the carbon source. In the morning, 500 µl of an overnight culture was used to inoculate 50 ml of minimal medium with benzoate or succinate as the sole carbon source. Cell growth was monitored turbidometrically with a Klett-Summerson colorimeter.
**TABLE 2.1. A. baylyi strains and plasmids**

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant characteristics&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Reference or Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. baylyi strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADP1</td>
<td>Wild type (BD413)</td>
<td>(18)</td>
</tr>
<tr>
<td>ISA36</td>
<td>benM::ΩS4036</td>
<td>(12)</td>
</tr>
<tr>
<td></td>
<td>catM3102&lt;sup&gt;b&lt;/sup&gt;</td>
<td>(23)</td>
</tr>
<tr>
<td>ADP102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACN47</td>
<td>benM::ΩS4036 benA::lacZ-Km&lt;sup&gt;i&lt;/sup&gt;5032</td>
<td>(12)</td>
</tr>
<tr>
<td>ACN146</td>
<td>benM::ΩS4036 benMA5146</td>
<td>(12)</td>
</tr>
<tr>
<td>ACN153</td>
<td>benM::ΩS4036 catM5153 (CatM[V158M])</td>
<td>This study</td>
</tr>
<tr>
<td>ACN157</td>
<td>benA::lacZ-Km&lt;sup&gt;i&lt;/sup&gt;5032 benM::ΩS4036 benMA5146</td>
<td>(12)</td>
</tr>
<tr>
<td>ACN164</td>
<td>benM::ΩS4036 benA::lacZ-Km&lt;sup&gt;i&lt;/sup&gt;5032 catM5153 (CatM[V158M])</td>
<td>This study</td>
</tr>
<tr>
<td>ACN293</td>
<td>benM::ΩS4036 benMA5147 ΔcatM5293</td>
<td>(27)</td>
</tr>
<tr>
<td>ACN539</td>
<td>benM::ΩS4036 benMA5147 ΔcatM5293 catB::lacZ-Km&lt;sup&gt;i&lt;/sup&gt;5534</td>
<td>This study</td>
</tr>
<tr>
<td>ACN541</td>
<td>benM::ΩS4036 benMA5146 catMΩK5541</td>
<td>This study</td>
</tr>
<tr>
<td>ACN547</td>
<td>benM::ΩS4036 benMA5146 catM3102 (CatM[R156H])</td>
<td>This study</td>
</tr>
<tr>
<td>ACN548</td>
<td>benM::ΩS4036 benMA5146 benA::lacZ-Km&lt;sup&gt;i&lt;/sup&gt;5032 catM3102 (CatM[R156H])</td>
<td>This study</td>
</tr>
<tr>
<td>ACN549</td>
<td>catM3102 (CatM[R156H])</td>
<td>This study</td>
</tr>
<tr>
<td>ACN558</td>
<td>benM::ΩS4036 catM3102 (CatM[R156H])</td>
<td>This study</td>
</tr>
<tr>
<td>ACN559</td>
<td>benM::ΩS4036 benA::lacZ-Km&lt;sup&gt;i&lt;/sup&gt;5032 catM3102 (CatM[R156H])</td>
<td>This study</td>
</tr>
<tr>
<td>ACN560</td>
<td>benM::ΩS4036 catB::lacZ-Km&lt;sup&gt;i&lt;/sup&gt;5534 catM5153 (CatM[V158M])</td>
<td>This study</td>
</tr>
<tr>
<td>ACN561</td>
<td>benM::ΩS4036 benMA5146 catB::lacZ-Km&lt;sup&gt;i&lt;/sup&gt;5534 catM3102 (CatM[R156H])</td>
<td>This study</td>
</tr>
<tr>
<td>ACN585</td>
<td>benM::ΩS4036 catB::lacZ-Km&lt;sup&gt;i&lt;/sup&gt;5534</td>
<td>This study</td>
</tr>
<tr>
<td>Plasmids</td>
<td>Description</td>
<td>Source</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>PUC13, pUC19</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, cloning vector</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, cloning vector</td>
</tr>
<tr>
<td>pRK415</td>
<td>Tc&lt;sup&gt;i&lt;/sup&gt;, broad host range cloning vector</td>
<td>Tc&lt;sup&gt;i&lt;/sup&gt;, broad host range cloning vector</td>
</tr>
<tr>
<td>pCR2.1-TOPO</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, PCR cloning vector</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pET-21b</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, T7 expression vector</td>
<td>Novagen</td>
</tr>
<tr>
<td>pHIP45</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, Sm&lt;sup&gt;i&lt;/sup&gt;Sp&lt;sup&gt;i&lt;/sup&gt;, source of Ω&lt;sup&gt;S&lt;/sup&gt;</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, Sm&lt;sup&gt;i&lt;/sup&gt;Sp&lt;sup&gt;i&lt;/sup&gt;, source of Ω&lt;sup&gt;S&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKOK6</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Km&lt;sup&gt;i&lt;/sup&gt;, Source of promoterless lacZ-Km&lt;sup&gt;i&lt;/sup&gt; cassette</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Km&lt;sup&gt;i&lt;/sup&gt;, Source of promoterless lacZ-Km&lt;sup&gt;i&lt;/sup&gt; cassette</td>
</tr>
<tr>
<td>pUII1637</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Km&lt;sup&gt;i&lt;/sup&gt;, source of Ω&lt;sup&gt;K&lt;/sup&gt;</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Km&lt;sup&gt;i&lt;/sup&gt;, source of Ω&lt;sup&gt;K&lt;/sup&gt;</td>
</tr>
<tr>
<td>pIB17</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; catM&lt;sup&gt;3102&lt;/sup&gt; (11950 – 13205)&lt;sup&gt;c&lt;/sup&gt; in pUC19</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; catM&lt;sup&gt;3102&lt;/sup&gt; (11950 – 13205)&lt;sup&gt;c&lt;/sup&gt; in pUC19</td>
</tr>
<tr>
<td>PIGG4</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; catM fragment (11950 – 12892)&lt;sup&gt;c&lt;/sup&gt; in pUC19</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; catM fragment (11950 – 12892)&lt;sup&gt;c&lt;/sup&gt; in pUC19</td>
</tr>
<tr>
<td>pIGG14</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Sm&lt;sup&gt;i&lt;/sup&gt;Sp&lt;sup&gt;i&lt;/sup&gt;, catBCIJFD (12892 – 18153)&lt;sup&gt;c&lt;/sup&gt; in pUC13 with Ω&lt;sup&gt;S&lt;/sup&gt; inserted in EcoRV site (15660)&lt;sup&gt;c&lt;/sup&gt; of cat&lt;sup&gt;J&lt;/sup&gt;</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Sm&lt;sup&gt;i&lt;/sup&gt;Sp&lt;sup&gt;i&lt;/sup&gt;, catBCIJFD (12892 – 18153)&lt;sup&gt;c&lt;/sup&gt; in pUC13 with Ω&lt;sup&gt;S&lt;/sup&gt; inserted in EcoRV site (15660)&lt;sup&gt;c&lt;/sup&gt; of cat&lt;sup&gt;J&lt;/sup&gt;</td>
</tr>
<tr>
<td>pBAC6</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Km&lt;sup&gt;i&lt;/sup&gt;, fragment (11950 – 12892)&lt;sup&gt;c&lt;/sup&gt; containing catM in pUC19 (Ω&lt;sup&gt;K&lt;/sup&gt; in HincII site [12687]&lt;sup&gt;c&lt;/sup&gt; of catM)</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Km&lt;sup&gt;i&lt;/sup&gt;, fragment (11950 – 12892)&lt;sup&gt;c&lt;/sup&gt; containing catM in pUC19 (Ω&lt;sup&gt;K&lt;/sup&gt; in HincII site [12687]&lt;sup&gt;c&lt;/sup&gt; of catM)</td>
</tr>
<tr>
<td>pBAC44</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Sm&lt;sup&gt;i&lt;/sup&gt;Sp&lt;sup&gt;i&lt;/sup&gt;, benMABCDE (563–7876)&lt;sup&gt;c&lt;/sup&gt; in pUC19, (Ω&lt;sup&gt;S&lt;/sup&gt; in multiple cloning site of vector)</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Sm&lt;sup&gt;i&lt;/sup&gt;Sp&lt;sup&gt;i&lt;/sup&gt;, benMABCDE (563–7876)&lt;sup&gt;c&lt;/sup&gt; in pUC19, (Ω&lt;sup&gt;S&lt;/sup&gt; in multiple cloning site of vector)</td>
</tr>
<tr>
<td>pBAC54</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Km&lt;sup&gt;i&lt;/sup&gt;, lacZ-Km&lt;sup&gt;i&lt;/sup&gt; cassette in NsiI site (3761)&lt;sup&gt;c&lt;/sup&gt; in ben&lt;sup&gt;A&lt;/sup&gt; with adjacent ben region (2316 – 5663)&lt;sup&gt;c&lt;/sup&gt; in pUC19</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt; Km&lt;sup&gt;i&lt;/sup&gt;, lacZ-Km&lt;sup&gt;i&lt;/sup&gt; cassette in NsiI site (3761)&lt;sup&gt;c&lt;/sup&gt; in ben&lt;sup&gt;A&lt;/sup&gt; with adjacent ben region (2316 – 5663)&lt;sup&gt;c&lt;/sup&gt; in pUC19</td>
</tr>
<tr>
<td>pBAC200</td>
<td>Tc&lt;sup&gt;i&lt;/sup&gt;, cat-region DNA (9819 to 10649&lt;sup&gt;c&lt;/sup&gt; and 15951 to 18153&lt;sup&gt;c&lt;/sup&gt;) in pRK415. Used to isolate the chromosomal cat&lt;sup&gt;MB&lt;/sup&gt; region</td>
<td>Tc&lt;sup&gt;i&lt;/sup&gt;, cat-region DNA (9819 to 10649&lt;sup&gt;c&lt;/sup&gt; and 15951 to 18153&lt;sup&gt;c&lt;/sup&gt;) in pRK415, used to isolate the chromosomal cat&lt;sup&gt;MB&lt;/sup&gt; region</td>
</tr>
<tr>
<td>pBAC234</td>
<td>Tc&lt;sup&gt;i&lt;/sup&gt;, ACN153 cat-region DNA (9819-18153&lt;sup&gt;c&lt;/sup&gt;) in pRK415, Isolated with pBAC200, cat&lt;sup&gt;MB&lt;/sup&gt; region (563-7876)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Tc&lt;sup&gt;i&lt;/sup&gt;, ACN153 cat-region DNA (9819-18153&lt;sup&gt;c&lt;/sup&gt;) in pRK415, Isolated with pBAC200, cat&lt;sup&gt;MB&lt;/sup&gt; region (563-7876)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pBAC284</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, catM segment (11950-12892)&lt;sup&gt;c&lt;/sup&gt; from pBAC234 cloned in pUC19, cat&lt;sup&gt;MB&lt;/sup&gt; region (563-7876)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, catM segment (11950-12892)&lt;sup&gt;c&lt;/sup&gt; from pBAC234 cloned in pUC19, cat&lt;sup&gt;MB&lt;/sup&gt; region (563-7876)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pBAC364</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, benMA5146 intergenic region (2293-2540)&lt;sup&gt;d&lt;/sup&gt;&lt;sup&gt;c&lt;/sup&gt; in pUC19, DNase I footprinting, antisense strand labeling</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, benMA5146 intergenic region (2293-2540)&lt;sup&gt;d&lt;/sup&gt;&lt;sup&gt;c&lt;/sup&gt; in pUC19, DNase I footprinting, antisense strand labeling</td>
</tr>
<tr>
<td>pBAC366</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, wild-type benMA intergenic region (2293-2540)&lt;sup&gt;d&lt;/sup&gt;&lt;sup&gt;c&lt;/sup&gt; in pUC19, DNase I footprinting, antisense strand labeling</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, wild-type benMA intergenic region (2293-2540)&lt;sup&gt;d&lt;/sup&gt;&lt;sup&gt;c&lt;/sup&gt; in pUC19, DNase I footprinting, antisense strand labeling</td>
</tr>
<tr>
<td>pBAC371</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, benMA5146 region (2316-2540)&lt;sup&gt;d&lt;/sup&gt;&lt;sup&gt;c&lt;/sup&gt;, HindIII deletion of</td>
<td>Ap&lt;sup&gt;i&lt;/sup&gt;, benMA5146 region (2316-2540)&lt;sup&gt;d&lt;/sup&gt;&lt;sup&gt;c&lt;/sup&gt;, HindIII deletion of</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Description</td>
<td>Source</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>pBAC364</td>
<td>DNase I footprinting, sense strand labeling</td>
<td></td>
</tr>
<tr>
<td>pBAC373</td>
<td>Ap&lt;sup&gt;+&lt;/sup&gt;, wild-type <em>benMA</em> region (2316-2540)&lt;sup&gt;d&lt;/sup&gt;&lt;sup&gt;c&lt;/sup&gt; <em>HindIII</em> deletion of pBAC366, DNase I footprinting, sense strand labeling</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC383</td>
<td>Ap&lt;sup&gt;+&lt;/sup&gt;, PCR fragment with <em>catM5153</em> in the pET-21b, for purification of CatM[V158M]</td>
<td></td>
</tr>
<tr>
<td>pBAC669</td>
<td>Ap&lt;sup&gt;+&lt;/sup&gt;, <em>catB</em>&lt;sup&gt;CIJ&lt;/sup&gt; (12892-15658)&lt;sup&gt;c&lt;/sup&gt; in pUC13</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC673</td>
<td>Ap&lt;sup&gt;+&lt;/sup&gt;, <em>catB</em> (13205-14225)&lt;sup&gt;c&lt;/sup&gt; in pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC674</td>
<td>Ap&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;, <em>catB</em> (13205-14225)&lt;sup&gt;c&lt;/sup&gt; upstream of <em>lacZ</em>-Km&lt;sup&gt;+&lt;/sup&gt; cassette in <em>SalI</em> site (14225)&lt;sup&gt;c&lt;/sup&gt; of <em>catB</em> in pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC675</td>
<td>Ap&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;, <em>catB</em> (13205-14225)&lt;sup&gt;c&lt;/sup&gt; <em>lacZ</em>-Km&lt;sup&gt;+&lt;/sup&gt; <em>catJFD</em> (15660–17347)&lt;sup&gt;c&lt;/sup&gt; in pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC679</td>
<td>Ap&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;, PCR fragment with <em>catM3102</em> in pCR2.1-TOPO</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC 684</td>
<td>Ap&lt;sup&gt;+&lt;/sup&gt;, 0.91kb *catM3102 NdeI&lt;sup&gt;d&lt;/sup&gt;-<em>XhoI&lt;sup&gt;d&lt;/sup&gt;</em> fragment from pBAC679 in pET-21b, complete coding sequence</td>
<td>This study</td>
</tr>
</tbody>
</table>

<sup>a</sup>Ap<sup>+</sup>, ampicillin resistant; Te<sup>-</sup>, tetracycline resistant; Sm<sup>-</sup>, streptomycin resistant; Sp<sup>-</sup>, spectinomycin resistant; Km<sup>-</sup>, kanamycin resistant; Ω<sub>S</sub>, omega cassette conferring Sm<sup>-</sup>Sp<sup>-</sup>; Ω<sub>K</sub>, omega cassette conferring Km<sup>-</sup>

<sup>b</sup>The original strain isolated with the *catM3102* allele has a mucoid colony morphology and is likely to carry additional uncharacterized mutations.

<sup>c</sup>Position in the *ben-cat* sequence in GenBank entry (accession number AF009224)

<sup>d</sup>Restriction sites added in primers used to generate PCR-amplified fragment for cloning
DNA sequencing and plasmid construction

Standard methods were used for DNA purification, digestion, ligation, electrophoresis, and bacterial transformation (30). Sequencing was carried out at the University of Georgia Integrated Biotech Laboratories. Plasmids are listed in Table 2.1.

To purify CatM[V158M], pBAC383, was made as was CatM-encoding pBAC381 (8) except the catM5153 allele was the template for PCR amplification. To generate a catB::lacZ fusion, the promoterless lacZ-Km\textsuperscript{r} cartridge of pKOK6 (20) was inserted into the Sal\textsuperscript{I} site in catB of pBAC673 to form pBAC674. To enable allelic exchange with the chromosome, cat-region DNA was inserted downstream of the lacZ-Km\textsuperscript{r} cartridge as follows. A catJFD fragment was isolated as a BamHI-Asp718 fragment from pIGG14 and cloned into pBAC674 digested with BamHI and Asp718 to form pBAC675. Plasmid pBAC679 was constructed in which the catM3102 allele was PCR amplified from pIB17 template and cloned into pCR2.1-TOPO. At sites introduced by the amplification primers, the catM3102 allele fragment was excised from this plasmid by digestion with Nde\textsuperscript{I} and Xho\textsuperscript{I} and ligated into similarly digested pET21b (Novagen) to create pBAC684. A plasmid with a catM disruption, pBAC6, was constructed by inserting an omega cassette encoding kanamycin resistance (14) into the HincII site of pIGG4.

BenM-independent Ben\textsuperscript{+} mutants

Spontaneous mutants of benM-disrupted ISA36 (Table 2.1) that grow on benzoate (Ben\textsuperscript{+}) were selected after incubation on solid benzoate medium (12, 13). The catM-catB regions from Ben\textsuperscript{+} mutants were isolated by gap-repair (17). DNA segments were tested for the ability to transform ISA36 to a Ben\textsuperscript{+} phenotype (12, 13). The cat-region of
ACN153 was isolated on pBAC234. The mutation conferring Ben+ growth was localized to the catM segment on pBAC284, which was subjected to DNA sequence analysis.

**Generation of *A. baylyi* strains by allelic exchange**

Plasmid-borne alleles were introduced into the chromosome by methods that exploit the high efficiency of natural transformation and recombination in ADP1-derived strains (12, 23). Briefly, recipients are transformed with DNA, typically linearized plasmids or crude cell-free lysates. Transformants in which homologous recombination has replaced the corresponding chromosomal region of the recipient with the donor DNA are initially identified by phenotypic changes. Strains generated for this study were tested for antibiotic resistances and carbon source utilization. Strains with the catM3102 allele were tested for the characteristic high CatA enzyme activity in succinate-grown cultures (12, 23). Genotypes were confirmed by Southern hybridization, analysis of PCR-generated fragment sizes and/or DNA sequencing of chromosomal regions.

With these methods, ACN153 was transformed with pBAC54 digested with Asp718. In a transformant, ACN164, homologous recombination had replaced the chromosomal benA locus with the fragment-borne benA::lacZ fusion. Similarly, a DNA fragment with catB::lacZ was made by digesting pBAC675 with Asp718. This fragment replaced the chromosomal catB-F region of ISA36 and ACN153 to make ACN585 and ACN560, respectively. With the exception that it has the catB::lacZ fusion, ACN539 is isogenic to ACN293.

Several strains were constructed in multiple steps. To facilitate the chromosomal introduction of the catM3102 allele, a drug resistance marker was first inserted in catM.
Exchange of the catM alleles could then be assessed by the acquisition of drug sensitivity. Plasmid pBAC6, digested with XmnI, was used to introduce the drug resistance marker into the chromosome of ACN146 to generate ACN541. A DNA fragment carrying the catM3102 allele (pBAC684 digested with XhoI) was used to transform ACN541. In strain ACN547, the catM3102 allele replaced the marker-disrupted catMΩK5541 of the recipient strain, ACN541. Strain ACN547 served as the recipient when transformed by a DNA fragment carrying the catB::lacZ fusion (pBAC675 digested with Asp718). Introduction of this lacZ reporter into the chromosome by allelic exchange yielded strain ACN561. Strain ACN547 was the recipient in a transformation with donor DNA from ACN157 (in the form of a cell-free lysate) used to introduce the benA::lacZ fusion into the chromosome. The resulting strain was ACN548.

ACN559 was also made in steps. ACN547 was transformed with wild-type benM-benA DNA (pBAC44 digested with XmnI). DNA sequencing confirmed that a transformant, ACN549, had wild-type ben region DNA. Next, the wild-type benM of ACN549 was replaced with the benM-disrupted allele of ISA36 by transforming the former strain with a cell-free lysate of the latter to generate ACN558. In the final step, a cell-free lysate of ACN47 was used to introduce the benA::lacZ fusion into the chromosome of ACN558 to generate ACN559.

**β-galactosidase (LacZ) assays**

To assay the benA::lacZ transcriptional fusion, cultures were grown in minimal medium with muconate or succinate as the carbon source. To assay the catB::lacZ transcriptional fusion, cultures were grown in minimal medium on succinate. To assess
induction, muconate (1 mM final concentration) was added to the growth medium for half the cultures. Growth was monitored by optical density (OD₆₀₀), and assays were done when cultures reached stationary phase. Culture samples (2 to 20 µl per assay) were lysed with sodium dodecyl sulfate and chloroform. Assay directions were followed for the FlourAce β-galactosidase reporter kit (Bio-Rad). The product of substrate hydrolysis, 4-methylumbelliferone (4MU), was detected with a TD-360 mini-fluorometer (Turner Designs). Relative fluorescence unit measurements enabled 4MU quantification by comparison with a standard curve.

**Purification of CatM[V158M]**

A 50 ml culture of BL21(DE3)(pBAC383) was grown overnight in LB with ampicillin and used to inoculate 1L of the same medium. After 4 h incubation at 37 °C, on a shaking platform, isopropyl-β-D-thiogalactopyranoside (1 mM final concentration) was added as inducer. Following further incubation for 4 h, cells were harvested by centrifugation (6,000 x g). The cell pellet was stored at -70 °C. The pellet was suspended in 30 ml buffer A (50 mM Tris[pH 8], 5 mM dithiothreitol, 10% [vol/vol] glycerol), and a crude extract was prepared by sonication. Following centrifugation (6,000 x g), the supernatant fraction was filtered through a 0.22 µm syringe filter, and subsequent purification was carried out with an FPLC system (Pharmacia). The CatM[V158M] protein failed to bind to a 5 ml Hi-Trap heparin column (Pharmacia) and was recovered from the flow-through fraction after 35% ammonium sulfate precipitation. The precipitate, which was collected by centrifugation (12,000 x g), was suspended in 7 ml buffer A and passed through a 5 ml Hi-Trap desalting column (Pharmacia). The protein
sample was next loaded onto a 5 ml Hi-Trap heparin column that was then washed with 75 ml buffer A. Protein was eluted with 75 ml of a 0 to 0.6 M NaCl gradient in buffer A. Fractions containing CatM[V158M] were pooled after identification by sodium dodecyl sulfate polyacrylamide gel electrophoresis. Using gels stained with Coomassie brilliant blue R-250 (BioRad), the CatM[V158M] protein was estimated to be > 90% pure. The Bradford method was used to assay protein concentrations with bovine serum albumin as the standard (3).

**DNase I footprinting**

As previously described, pBAC366 and pBAC373 were used to generate antisense and sense fragments of the wild-type *benMA* region end labeled with [γ-32P] (6). Similarly, pBAC364 and pBAC371 were used to produce labeled fragments of the *benMA5146* region. For use in footprint reactions, the CatM and BenM proteins were purified by cation exchange and heparin-agarose chromatography as described elsewhere (6, 8). Various concentrations of BenM, CatM, or CatM[V158M] were incubated with the radiolabeled DNA probe (circa 300 pM; 200,000 cpm) at 30 °C for 30 min with or without benzoate or muconate (0 to 5 mM). The binding reaction (20 µl total volume) was carried out in the following buffer: 80 mM Tris-acetate (pH 8.0), 100 mM potassium acetate, 25 mM ammonium acetate, 5 mM magnesium acetate, 0.1 mM EDTA, 1.0 mM dithiothreitol, 1 mM calcium chloride, 2 µg/ml calf thymus DNA, 50 µg/ml bovine serum albumin. DNase I cleavage reactions were carried out for 1 min at 30 °C and analyzed with the methods used by Wang and Hoover (34).
RESULTS

A catM point mutation increases expression of benA without BenM

Sequence similarity between benM and catM suggested that CatM variants might substitute for BenM. To test this possibility, spontaneous mutants were isolated that grow on benzoate without BenM. Ten independent mutants derived from a benM-disrupted parent strain (ISA36) were selected on benzoate medium, as previously described (12, 13). From each mutant, chromosomal DNA in the catM region was isolated and tested for its ability to transform the parent strain to grow on benzoate. In one mutant (ACN153), the coding sequence of catM was responsible for conferring this Ben+ phenotype. Sequence analysis of the mutant allele, designated catM5153, identified a single mutation, a G to A transition at position 474 of catM with respect to its translational start site. In the deduced sequence of the variant protein, CatM[V158M], a methionine residue substitutes for valine at position 158 of the 303-residue protein.

Since the primary obstacle to BenM-independent growth on benzoate is low expression of the benABCDE operon (12, 13), the effect of the catM mutation on ben-gene expression was assessed. A benA::lacZ transcriptional fusion was introduced into the chromosome of ACN153, which carries the catM mutation. In the resulting ACN164, the replacement of the wild-type benA allele by the transcriptional fusion prevents consumption of benzoate as sole carbon source. This strain was grown on muconate or succinate as the carbon source, and activity of the lacZ-encoded enzyme, β-galactosidase, was evaluated in relationship to that of the comparable strain with a wild-type catM allele (ACN47). The benM gene is disrupted in all strains tested for expression of the benA::lacZ fusion. With either the wild-type or mutant catM allele, benA::lacZ expression
was inducible by growth on muconate (Fig. 2.2). However, the mutant allele caused an approximately two-fold increase in expression relative to that in ACN47. To test whether the variant CatM might recognize benzoate as an effector, lacZ activity in ACN164 was measured with benzoate added to the growth medium alone or in combination with muconate. However, no increase in gene expression was observed in response to benzoate (data not shown) (11).

**Effect of the catM5153 allele on catB expression**

It was not clear whether the 2-fold change in muconate-inducible ben-gene expression was responsible for the ability of ACN153, which has catM5153-encoded CatM[V158M], to form colonies on benzoate medium within three days. The comparable strain with wild-type catM does not form colonies on such medium even after prolonged incubation. We tested whether catM5153 might additionally affect catB expression. Regulation of this gene was studied because CatM-dependent benA expression is augmented by decreased CatB enzyme activity (13). This augmentation may result from the transient accumulation of muconate, the co-activator of CatM, when there are reduced levels of CatB, the enzyme that degrades muconate (Fig. 2.1) (13). Effects related to internal muconate concentration during growth on benzoate would not be observed under our assay conditions since the chromosomal benA::lacZ fusion prevents the conversion of benzoate to muconate (Figs. 2.1-2). To test the effect of CatM[V158M] on catB expression, a catB::lacZ transcriptional fusion was used. Activity of LacZ was compared between ACN585, with the wild-type catM, and ACN560, with catM5153. In both strains benM is disrupted. The fusion replaces catB on the chromosome, thereby preventing
FIG. 2.2. Expression of a chromosomal *benA::lacZ* fusion in strains encoding CatM or variant regulators (CatM[V158M], encoded by *catM5153*, or CatM[R156H], encoded by *catM3102*). All strains lack a functional *benM*, and two strains have a point mutation at position –40 relative to the *benA* transcription initiation site (*benMA5146*, see Fig. 2.1). Cultures were grown on succinate or muconate as indicated. β-Galactosidase (LacZ) activity is shown relative to that measured in succinate-grown ACN47 (6.9 μmol/min/ml/OD$_{600}$). Activities are the averages of at least four repetitions, and the standard deviations were <20% of the average value.
growth on muconate as the sole carbon source. Cultures were grown with succinate as the
carbon source in the presence or absence of muconate as an effector. Expression of the
catB::lacZ fusion was induced by muconate in both strains (Fig. 2.3). Deletion of catM in
a benM-disrupted strain (ACN539) eliminated the inducible β-galactosidase (LacZ)
activity (data not shown). This result supports the conclusion that CatM and
CatM[V158M] are the sole regulators of catB expression in ACN585 and ACN560,
respectively. Although catB expression in ACN560 increased in response to muconate,
the maximum level was approximately half that of ACN585. As noted previously,
lowered catB expression may increase the amount of muconate available to serve as the
CatM co-activator (13). Thus, during growth on benzoate, the catM5153 allele may
increase benA expression to an extent greater than that measured by the benA::lacZ fusion
(Fig. 2.2). Interestingly, the catM5153 effect on catB expression differed from that on
benA, a locus where the catM mutation caused expression levels to increase rather than
decrease relative to the wild-type catM (Fig. 2.2). These differences indicate that the
improved ability of CatM[V158M] to activate benA expression is not due to a general
increase in responsiveness to muconate. Nevertheless, the amino acid substitution at
residue 158 altered the response of CatM to muconate in the regulation of both benA and
catB. The position of this substitution was near that at residue 156 in a previously
characterized variant, CatM[R156H], that activates cat-gene transcription in the absence
of effectors (23).
FIG. 2.3. Expression of a chromosomal catB::lacZ fusion in strains encoding CatM or variant regulators (CatM[V158M], encoded by catM5153, or CatM[R156H], encoded by catM3102). All strains lack a functional benM. Cultures were grown on succinate with or without muconate as an inducer, as indicated. β-Galactosidase (LacZ) activity is shown relative to that measured in succinate-grown ACN585 (3.96 µmol/min/ml/OD<sub>600</sub>). Activities are the averages of at least three repetitions, and the standard deviations were <20% of the average value.
Relative β-Galactosidase Activity

- ACN585 CatM
- ACN560 CatM (V158M)
- ACN561 CatM (R156H)

- Succinate
- Succinate + Muconate

Activity levels:
- ACN585: 34
- ACN560 (V158M): 19
- ACN561 (R156H): 31
Effect of CatM[R156H] on benA expression

To improve our understanding of CatM-regulated benA expression, we re-examined catM3102, an allele with a point mutation encoding histidine rather than arginine at position 156 of CatM. This mutation was selected by its ability to confer high-level expression of the catIJF genes in the absence of muconate as the co-activator for CatM (23). When this catM mutation was isolated approximately twenty-five years ago, the benM and benA genes had not yet been identified in any organism.

The effect of this CatM[R156H] variant on benA expression had not previously been addressed. For this purpose we constructed ACN559, a benM-disrupted strain in which the catM3102 allele controls expression of the chromosomal benA::lacZ transcriptional fusion. As shown in Fig. 2.2, succinate-grown ACN559 expressed benA::lacZ at levels higher than those of ACN47, the comparable strain with wild-type catM. However, the catM mutation increased inducer-independent benA expression to a lesser extent than observed for catB. Studies of the catB::lacZ fusion indicated a 10-fold increase in expression for succinate-grown strain ACN561 compared to ACN585 with wild-type catM (Fig. 2.3). Nevertheless, the CatM[R156H] variant significantly altered benA expression relative to the wild-type CatM. In muconate-grown cells, the benA::lacZ expression level was approximately 5-fold higher in ACN559 than ACN47 (Fig. 2.2). This high level of benA expression in ACN559 raised the possibility that CatM[R156H] might substitute for BenM during growth on benzoate.
**BenM-independent growth on benzoate**

To test the effect of *catM3102* on benzoate catabolism, strain ACN558 was engineered to contain the wild-type *benA* region, the disrupted *benM* allele, and the *catM* mutation encoding CatM[R156H] (Table 2.1). ACN558 was able to grow with benzoate as the sole carbon source, although it grew slower than the wild type (ADP1) with both a longer generation time and a longer lag period (Table 2.2). This pattern of growth was nearly identical to that of ACN153, the strain with the CatM[V158M] variant (Table 2.2). These results demonstrate that a single CatM regulatory protein can control expression of both the *cat* and *ben* genes at levels sufficient to permit benzoate consumption without BenM.

To study BenM-independent regulation further, we used a previously isolated mutation that increases CatM-activated *benA* transcription (*benMA5146*, with a T to A transversion at –40 relative to the *benA* transcript start, Fig. 2.1C) (12). The effect of this mutation on *benA::lacZ* expression was studied in succinate- and muconate-grown cultures (Fig. 2.2, strain ACN157). The mutation increased the level of muconate-inducible *benA* expression relative to the wild-type promoter (in ACN47). CatM-regulated *benA* expression patterns were similar to those of the CatM[R156H] variant at the wild-type *benA* promoter (strain ACN559, Fig. 2.2). Thus, a single mutation in either *catM* or the *benA* promoter is sufficient to increase CatM-activated *ben*-gene expression. This *benMA5146* promoter mutation in ACN146 did not noticeably affect growth. With benzoate as the carbon source, the growth curves for ACN146 and the wild type were nearly the same (Table 2.2). Therefore, mutations that allow CatM to regulate BenM-independent growth on benzoate do not necessarily cause slower growth rates.
### TABLE 2.2. Effect of *catM* and *benA* Mutations on rates of BenM-independent Growth on Benzoate$^a$

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant Characteristics $^b$</th>
<th>Generation Time (min)$^c$</th>
<th>Lag Time (h)$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADP1</td>
<td>Wild Type</td>
<td>60 ± 10</td>
<td>7 ± 0.5</td>
</tr>
<tr>
<td>ISA36</td>
<td>No BenM, wild-type CatM</td>
<td>No Growth</td>
<td>No Growth</td>
</tr>
<tr>
<td>ACN558</td>
<td>No BenM, CatM[R156H]</td>
<td>96 ± 8</td>
<td>9 ± 0.5</td>
</tr>
<tr>
<td>ACN153</td>
<td>No BenM, CatM[V158M]</td>
<td>93 ± 4</td>
<td>10 ± 0.5</td>
</tr>
<tr>
<td>ACN146</td>
<td>No BenM, -40 <em>benA</em> promoter mutation, wild-type CatM</td>
<td>50 ± 11</td>
<td>7 ± 0.5</td>
</tr>
<tr>
<td>ACN547</td>
<td>No BenM, -40 <em>benA</em> promoter mutation, CatM[R156H]</td>
<td>59 ± 8</td>
<td>4 ± 0.5</td>
</tr>
</tbody>
</table>

$^a$Provided as the sole carbon source. When succinate was provided as the carbon source, the growth rates of all these strains were comparable (data not shown).

$^b$See Table 2.1 for genotypes

$^c$Averages of three or more determinations

$^d$Time between inoculation and start of exponential growth
**Combined effects of point mutations in the benA promoter and the catM gene**

We tested the effect of the *benMA5146* promoter mutation on the ability of CatM[R156H] to activate *benA* expression. In ACN548, which lacks BenM and carries the *catM3102* allele, the *benA::lacZ* chromosomal fusion is under the control of the *benMA5146* promoter. Interestingly, the combination of the promoter mutation and the *catM* mutation resulted in significantly increased levels of *benA* expression in succinate-grown cells (Fig. 2.2 ACN548). Expression of the *benA::lacZ* fusion in this strain was further induced by growth on muconate. This combination of two mutations enabled a *benM* mutant, ACN547, to grow on benzoate with a generation time comparable to that of the wild type yet with a shorter lag time than the other strains in Table 2.2. Rapid initiation of growth may reflect increased muconate-independent expression of the *benABCDE* operon (as indicated by succinate-grown ACN548, Fig. 2.2). High levels of *ben*-gene expression prior to the metabolic formation of muconate may allow benzoate consumption to commence without delay. Moreover, the ability of the *benMA5146* promoter mutation to alter CatM-regulated *benA* expression with both the wild-type and variant regulators raised questions about promoter-protein interactions.

**CatM interactions with the benA promoter region**

The N-terminal 58 amino acids of BenM are 85% identical to those of CatM. This resemblance predicts nearly identical DNA binding domains (22, 31, 32). To test whether BenM and CatM bind similarly to the *benA* promoter, DNase I footprinting was used. BenM, in the absence of effectors, binds to two regions upstream of the *benA* transcriptional start site (6). One region, designated Site 3, overlaps the –10 region of the
*benA* promoter. The second region, designated Site 1, is adjacent to the –35 region of the *benA* promoter (Fig. 2.1C and 2.4E). Here, experiments with BenM (Fig. 2.4A and C, lane 2) were used as comparisons for those with CatM. Similar to BenM in the absence of effectors, the wild-type CatM (Fig. 2.4B lane 2 and 2.4D lane 3) or the CatM[V158M] variant (Fig. 2.4D, lane 2), protected the Site 1 and 3 regions from DNase I cleavage as compared to a reaction without any regulatory protein (lane 1 in each panel). The binding of CatM or CatM[V158M] to the *benA* region is consistent with a previous model for BenM in which a regulatory tetramer, in the absence of effectors, represses *benA* transcription (top of Fig. 2.5).

With no effectors, BenM, CatM or CatM[V158M] rendered some positions between the protected regions hypersensitive to DNase I cleavage (triangles in Fig. 2.4E; sites labeled A-D in 2.4A and B). These regularly spaced cleavage sites, on both the sense and antisense strands, are separated by the distance of one DNA helical turn (10-11 nucleotides). One site on the antisense strand was more sensitive to DNase I cleavage in the presence of CatM relative to BenM (position –20, marked by an arrow labeled D in Fig. 2.4B and E). As interpreted for BenM, some sites may become more sensitive to DNase I cleavage than in the absence of protein via the formation of an exposed DNA loop when the tetrameric regulator binds. The spacing of the protein binding sites (1 and 3) places them on the same side of the DNA helix (6).

**Footprint changes in response to effectors**

The addition of effectors resulted in a loss of protection from DNase I cleavage in the Site 3 region (Fig. 2.4A-C, lanes 3-4; 4D, lanes 4-5). Benzoate did not affect the DNase I
**FIG. 2.4.** DNase I footprinting of BenM, CatM and CatM[V158M] at *benA*. DNase I-cleaved DNA was labeled on the antisense (A and B) or sense strand (C and D) of the *benA* promoter region. In lanes 2-5, a regulatory protein (0.15 µM) was present in the cleavage reaction as indicated below panels A-D. The presence (+) or absence (-) of the effectors benzoate (Ben) and muconate (Muc) at a concentration of 1 mM in each reaction is shown. The binding sites labeled 1-3 are discussed in the text, and their positions are shown relative to the DNA sequence of the region (E). Nucleotides protected from DNase I digestion in both the absence and presence of inducers are indicated by solid brackets immediately above or below the *benA* sequence. Nucleotides protected from DNase I cleavage only in the absence of inducers are indicated by dotted brackets. Triangles and positions labeled A-D show sites that were hypersensitive to DNase I digestion. One hypersensitive site (D) was evident in reactions with CatM but not BenM.
**FIG. 2.5.** Model of regulated *benA* expression. Footprint data (Fig. 2.4) suggest that CatM binds to the same regions of the *benA* promoter DNA as does BenM in the absence or presence of muconate. Models of *benA* regulation by BenM have been presented elsewhere (6, 9). BenM mediates higher levels of muconate-activated *benA* transcription than does CatM, as discussed in the text.
BenM or CatM

**Repressed transcription**

- RNA polymerase
- Site 1
- Site 2
- Site 3
- +1
- benA

**Activated transcription**

- RNA polymerase
- Site 1
- Site 2
- Site 3
- +1
- benA

+ muconate

BenM or CatM
cleavage patterns in reactions containing wild type or variant CatM protein with or without muconate (data not shown), consistent with previous conclusions that CatM does not respond to benzoate (12). Loss of protection in the Site 3 region should improve the ability of RNA polymerase to access the –10 promoter region. These inducing conditions are associated with the transcriptional activation of *benA*. In the Site 2 region, the effectors altered the cleavage patterns. For example the BenM protein with both muconate and benzoate caused the continuous region of protection starting at Site 1 to extend in size on both strands (Fig. 2.4A and C, lane 4). Notably, the hypersensitive sites at –50 on the antisense strand and at –46 on the sense strand became protected. The cleavage sites at positions –39 and –29 on the antisense strand became much less enhanced. Comparable changes occurred in the corresponding regions on the sense strand. Transcriptional activation may occur when a regulatory tetramer binds to the Site 1 and Site 2 regions (Fig. 2.5, bottom).

With BenM and muconate alone (Fig. 2.4A and C, lane 3), the DNase I cleavage patterns were intermediate between those with no effectors or both effectors (Fig. 2.4A and C, lanes 2 and 4). Intermediate results also occur with benzoate as the sole effector (6). Similarly, this intermediate pattern resulted from wild-type CatM in the presence of muconate (Fig. 2.4B and D, lane 4). These conditions correspond to low-level *benA* transcription and may reflect a mixture of DNA fragments with protein bound to Sites 1 and 3 and others with protein bound to Sites 1 and 2. Increasing the muconate concentration up to 5 mM with CatM also increased protection of bands at –50 or –46 on the antisense or sense strands, respectively (data not shown) (11). This Site 2 position was also protected in the presence of muconate and CatM[V158M] (Fig. 2.4D, lane 5), an
arrangement that correlates with increased \textit{benA} transcription (Fig. 2.5, bottom). During \textit{benA} transcription, the Site 3 promoter region should interact with RNA polymerase and be unprotected in the absence of the polymerase. Thus, the anomalous Site 3 banding pattern with CatM[V158M] was surprising (Fig. 2.4D, lane 5). Occasionally, a similar result was observed in footprints with the wild-type BenM and CatM. Therefore, while the significance of this pattern is unclear, it may not depend solely on the variant CatM[V158M] protein.

The \textit{benMA5146} DNA region was also used in DNase I footprints (data not shown) (11). No major differences occurred in the positions to which CatM binds the DNA. In the absence or presence of muconate, the cleavage patterns for CatM and the \textit{benMA5146} region were similar to those for the wild-type \textit{benA} promoter with CatM or the CatM[V158M] variant (Fig. 2.4).

**DISCUSSION**

**CatM-	extit{benA} DNA interactions**

The overlapping DNA-binding functions of CatM and BenM reflect their sequence similarity, which is 85\% overall and 98\% in their N-terminal DNA binding domains. DNase I footprints suggested that without effectors BenM, CatM and the CatM[V158M] recognized LysR-type binding sequences within Site 1 and Site 3 regions of \textit{benA} DNA(Fig. 2.4A-D lanes 2 and 4D lane 3). Such interactions should repress basal \textit{benA} expression by blocking the –10 promoter region. Consistent with this interpretation, CatM-mediated repression occurs \textit{in vitro} and \textit{in vivo} in strains lacking BenM (6, 12).
Effector interactions with LysR-type proteins may cause global conformational changes in the tetramer (22). Effectors alter the number and/or the position of protein subunits bound to DNA (21, 31, 32). Effectors can also alter DNA bending and thereby impact transcription (1, 2). With benA DNA, muconate alleviated protection by BenM, CatM or CatM[V158M] in the –10 (Site 3) region of the promoter (Fig. 2.4A-C lanes 3, 4D lane 4). With the variant protein in this region, alleviation of protection was accompanied by an altered DNase I cleavage pattern of unknown significance (Fig. 2.4D lane 5 vs. 1). Muconate also affected the Site 2 region. With CatM, muconate reduced the extent of hypersensitivity in the cleavage of those positions indicated by triangles in Fig. 2.4E (Fig. 2.4B lane 3, 4D lane 4). These patterns with CatM and muconate in the Site 2 DNA were similar to those with BenM (Fig. 2.4A and 2.4C, lanes 3). The footprint patterns with muconate, which correlate with low-level benA transcription, may arise from a mixed population of DNA fragments in the active and repressed configurations (Fig. 2.5). Higher-level transcription correlates with protection of the cleavage site at position –50 on the antisense strand and –46 on the sense strand, as observed for BenM with muconate and benzoate (Fig. 2.4A and C, lanes 4) and CatM[V158M] with muconate (Fig. 2.4D lane 5).

Although there were some differences in the cleavage patterns, the overall benA regions to which the wild-type and variant CatM bound were similar to those of BenM. Therefore, the variations in benA expression levels could not be attributed to major shifts in the location of DNA-protein binding sites for BenM, CatM, CatM[V158M] and the mutant promoter benMA5146 (data not shown). Moreover, the extent of BenM and CatM similarity raised questions concerning the evolutionary retention of both paralogs.
Functional divergence of BenM and CatM

CatM served as the sole *ben-cat* regulator when spontaneous mutations occurred in *catM* (*catM*5153 and *catM*3102) or the *benA* promoter (*benMA*5146). Although the mutants rapidly consumed benzoate (Table 2.2), the relative fitness of the mutant and wild-type strains was not evaluated. Mutations that increased CatM-mediated *ben*-gene transcription altered the balance of *ben* and *cat* gene expression. CatM[V158M] and CatM[R156H] affected *benA* and *catB* differently (Figs. 2.2 and 2.3). During benzoate consumption (Fig. 2.1), the altered regulation of these genes should affect metabolite flow and the accumulation of muconate, a key effector that is also toxic at high concentrations (15).

This toxicity prevents growth on alternative carbon sources when muconate accumulates endogenously during the metabolism of aromatic precursors in a strain lacking *catB* (5, 35). Interestingly, exogenous muconate does not inhibit growth in the same fashion. Perhaps the coupled uptake and degradation of muconate prevent its intracellular accumulation (15). Further evidence that endogenous muconate is harmful comes from the analysis of spontaneous mutants. Those selected mutants that grow without *catB* in the presence of an appropriate aromatic precursor such as benzoate, anthranilate, benzyl alcohol or benzaldehyde, invariably acquire secondary mutations that block muconate formation (5, 35). By this method, numerous *catA* mutations have been obtained in which catechol accumulation is evident but not lethal (5, 35). Thus, the endogenous accumulation of muconate appears to be more toxic than that of catechol.

The slow growth of ACN558 and ACN153, which encode CatM variants, may reflect that these mutants are less adept than the wild type at balancing *ben* and *cat* gene
expression to optimize muconate concentrations during growth on benzoate (Table 2.2).

The key role of muconate as an effector allows this compound to control the genes needed for its own formation and degradation. The importance of muconate concentration is also suggested by the short lag time of ACN547, a strain with high inducer-independent expression of the *ben* genes (Table 2.2). In this strain, muconate should not need to accumulate to initiate pathway induction.

In the wild type, pathway induction is initiated by benzoate in conjunction with BenM. As assessed by transcriptional fusions (12), muconate alone causes BenM to activate higher *benA* expression levels than do the CatM variants (Fig. 2.2). Thus, muconate alone should be sufficient to induce benzoate consumption under laboratory conditions. Nevertheless, the intricate synergistic response of BenM to two effectors allows benzoate consumption to initiate quickly, to reach high levels in the presence of the substrate (benzoate) and a catabolite (muconate), and to decrease upon substrate depletion (6). During evolution, selection may favor regulatory schemes that optimize the ability to adapt to changing conditions (28). Furthermore, the use of two regulators could help balance *ben* and *cat* gene regulation when muconate is derived from substrates besides benzoate. Some aromatic compounds, such as anthranilate, are degraded via catechol such that *cat*- but not *ben*-gene expression is required (5, 15, 37). The complexity of the regulatory circuit provides the potential for very rapid and large variations in gene expression.
Promoter dependence of regulation

An intriguing aspect of the complexity is that CatM and BenM function differently at multiple promoters. Studies of the CatM[R156H] variant emphasize the importance of specific promoter sequences. This variant activates high-level muconate-independent catB transcription (Fig. 2.3). Surprisingly, the main effect of the CatM[R156H] variant on benA transcription was to increase inducer-dependent gene expression (Fig. 2.3 ACN559 versus ACN47). The different effects on catB and benA (Figs. 2.2 and 2.3) suggest that protein-DNA interactions in the benA region do not properly situate CatM for optimal RNA polymerase contact. It may be that the affinity of the benA promoter for CatM is reduced relative to that for BenM and/or that the conformations of the two regulators in their activated states are sufficiently different to affect transcription.

The benA and catB promoters have greater sequence differences between their Site 2 regions than the Site 1 and 3 regions (Fig. 2.6). Our regulatory model predicts that transcriptional activation results when a regulatory tetramer binds Site 1 and Site 2. The CatM[R156H] variant may have a higher proclivity for binding the catB Site 2 region in the absence of muconate than it has for the comparable region of benA. Consistent with this possibility, a mutation of T to A at position –40 in the Site 2 region of benA increased inducer-independent transcriptional activation by CatM[R156H] (Fig. 2.2, ACN548 versus ACN559). This mutation also increased CatM-activated inducible benA transcription (Fig. 2.2). In comparison to the wild-type sequence, the mutated sequence, AGTGT, more closely resembles the corresponding portion of the Site 1 region of catB,
FIG. 2.6. Comparison of sequences in the *benA* and *catB* promoter regions. The sequences are aligned relative to the transcription initiation site (+1), and identity is indicated (·). The significance of the *ben* region Sites 1, 2 and 3 is discussed in the text. Gray shading indicates a consensus sequence (ATAC-N7-GTAT) used to bind CatM (29), BenM (6) and other related members in a subclass of LysR-type regulators.
AGTAT, to which CatM binding has been demonstrated (Fig. 2.6) (29). It is not clear how the central portion of CatM affects protein binding to the benA Site 2 region. Nevertheless, mutations at positions 156 and 158, in the effector-binding domain of CatM, increase transcriptional activation of benA (Fig. 2.2). In other LysR-type regulators, mutations have been identified that affect both inducer and DNA binding. For example, amino acid substitutions in a central position of the NahR protein affect both inducer response and DNA binding (24). The effector-binding domains of diverse LysR-type regulators may share a common Rossmann-fold topology characteristic of a family of periplasmic binding proteins (26). This topology involves two domains connected by a hinge that allows movement when an effector binds in the inter-domain cavity. Residue 156 and 158 in CatM are predicted to be near this hinge-like region, and they could affect inter-domain movement. Alternatively, these positions might affect the oligomerization of CatM. When mutations important to transcriptional activation by LysR regulators were mapped relative to the structure of CbnR, many corresponded to residues at the interface between subunits (22). It is also possible that the two CatM substitutions affect transcription by different mechanisms. More information about effector-induced conformational changes in CatM and its variants should be provided by current structural studies (7, 9). Structural investigations, which may also reveal the basis for the inability of CatM to respond to benzoate, will complement the physiological characterization of the regulatory mutations in this report.
ACKNOWLEDGMENTS

We thank Nathaniel Cosper, Robert Scott, and Cory Momany for assistance with purification of the regulatory proteins. Timothy Hoover helped with the DNase I footprints and made many useful suggestions concerning the experiments and manuscript. We also thank Ann Onyenwoke and Matthew Wisdom who worked on this project as undergraduate students. This research was supported by National Science Foundation grants MCB-0212604 and MCB-0516914 to E.L.N. with an REU supplement for H.A.D.
REFERENCES


8. **Clark, T. J., C. Momany, and E. L. Neidle.** 2002. The *benPK* operon, proposed to play a role in transport, is part of a regulon for benzoate catabolism in *Acinetobacter* sp. strain ADP1. *Microbiology* **148**:1213-1223.


CHAPTER 3

DISTINCT EFFECTOR-BINDING SITES ENABLE SYNERGISTIC
TRANSCRIPTIONAL ACTIVATION BY BENM, A LYSR-TYPE REGULATOR

1Obidimma C. Ezezika, Sandra Haddad, Todd J. Clark, Ellen L. Neidle and Cory Momany. To be submitted to Cell.
O.C.E and S.H contributed equally to this report.
ABSTRACT

BenM, a bacterial transcriptional regulator, responds synergistically to two effectors, benzoate and cis,cis-muconate. CatM, a paralog with overlapping function, responds only to cis,cis-muconate. Structures of their effector-binding domains revealed two effector-binding sites in BenM. BenM and CatM are the first LysR-type regulators to be structurally characterized while bound to physiologically relevant inducers. The effector complexes were obtained by replacement of anions with cis,cis-muconate or benzoate. Crystals were soaked in liquors with high effector concentrations and low concentrations of competing ions. This strategy, including data collection with fragments of fractured crystals, may be generally applicable to related proteins. In BenM and CatM, the binding of cis,cis-muconate to an inter-domain pocket was facilitated by helix dipoles that provide charge stabilization. In BenM, benzoate also bound in an adjacent hydrophobic region where it alters the effect of cis,cis-muconate bound in the primary site. A charge relay system appears to underlie synergistic transcriptional activation.
INTRODUCTION

The BenM and CatM transcriptional regulators of the soil bacterium *Acinetobacter baylyi* ADP1 are paralogs with overlapping functions (8). BenM has the distinctive feature of activating transcription synergistically in response to two effectors, benzoate and *cis,cis*-muconate (hereafter designated muconate) (4). In contrast, CatM responds only to muconate. As reported here, comparisons of these two LysR-type transcriptional regulators (LTTRs), which are 59% identical in sequence, reveal the structural basis of their response to effectors. Despite the prevalence of the LTTR family (33), structural analyses of protein-effector interactions have been hampered by the inability to crystallize these regulators bound to their cognate small-molecule inducers.

LTTRs are the most common type of transcriptional regulator in proteobacteria (28). For example, strains of *Acinetobacter, Agrobacterium, Burkholderia, Escherichia, Pseudomonas,* and *Sinorhizobium* each have genomes predicted to encode approximately 40 to 120 family members. LTTRs regulate all types of metabolic function including amino acid biosynthesis, aromatic compound degradation, nodulation, oxidative stress, and virulence. The most conserved LTTR region is the N-terminal DNA-binding domain (33). In this domain, a winged-helix-turn-helix motif was confirmed in CbnR, the sole LTTR for which a full-length structure is known (23). The DNA-binding domain of CbnR connects to a two-domain regulatory region that resembles periplasmic binding proteins (31). This fold is conserved in the structures of several LTTR regulatory domains despite great sequence variability in this protein region (5, 34, 39).

The regulatory domain structures, when interpreted with respect to genetic studies, suggest that a hinge region between two sub-domains serves as an effector-
binding site. The binding of small molecular weight effectors most likely causes structural changes that alter DNA binding/bending and contact with RNA polymerase. Nevertheless, conformational changes involved in effector-mediated transcriptional activation by LTTRs remain unclear. In one regulator, OxyR, the structures were characterized of the inactive and active forms of the regulatory domain (5). However, this is a rare case where the LTTR does not bind an effector but instead responds to oxidation-state changes via disulfide bond formation.

Problems with low solubility make structural studies of LTTRs notoriously difficult. Successful X-ray crystallographic studies have relied on removal of the DNA-binding domain and the use of high-salt buffers. Such buffers create alternative problems by establishing competition for protein binding between ions in the crystallization buffer and the natural effectors. In studies of DntR and CysB, the presumed effector-binding sites were found to contain thiocyanate, acetate and/or sulfate ions that were inferred to mimic the natural ligands (34, 39, 42). Here, we report the direct visualization of biologically relevant ligands with the effector-binding domains (EBDs) of two LTTRs, BenM and CatM.

BenM and CatM control a complex regulon for aromatic compound degradation in *A. baylyi* ADP1 (Fig. 3.1). Transcriptional activation at four chromosomal loci enables benzoate consumption. BenM and CatM regulate each locus, although their relative importance varies by location. Additionally, during growth on benzoate, BenM and CatM repress genes used to consume alternative aromatic compounds (3). Both regulators respond to muconate, an effector that can cause activation or repression of gene expression. The functional overlap between BenM and CatM reflects a level of sequence
similarity that is 85% overall and 98% in the DNA binding domains. Nevertheless, the ability to respond to benzoate is unique to BenM. At the benA promoter, benzoate or muconate alone activates BenM-mediated transcription. Together, these compounds yield a BenM-dependent level of transcriptional activation that is higher than the sum of their individual effects (4, 10). This type of regulation is physiologically important. It helps balance the expression of multiple genes in a complex pathway and provides a rapid mechanism for integrating cellular signals (4, 9, 10). Here we present several structures of his-tagged BenM-EBD and CatM-EBD with and without effectors bound to the proteins. Models for the structural basis of transcriptional regulation and synergism are discussed.

MATERIALS AND METHODS

Chemicals

Reagent grade chemicals and Infinity Nanopure UF water at 18 M Ohm cm\(^{-1}\) were used to prepare solutions. The imidazole for protein purification, Fluka puris grade (>99.9%), had UV absorbent impurities evident at 280 nm. Muconate was purchased from Acros or provided as a gift from Celgene.

Histidine-tagged derivatives of CatM

To test whether the hexahistidine tag used for protein purification affected CatM function, an overlap-extension PCR method was used, as described for BenM (4). All Acinetobacter strains were derived from A. baylyi strain ADP1, a bacterium formerly designated Acinetobacter sp. or A. calcoaceticus (41). An allele was generated that could
FIG 3.1. Benzoate degradation in ADP1 is regulated by CatM and BenM and the effectors benzoate and muconate. The chromosomal *ben* and *cat* genes encode proteins for benzoate catabolism. BenM and CatM regulate transcription in the regions shown (black diamonds). In the intergenic *benMA* region, there are three potential binding sites for BenM and CatM, sites, 1, 2 and 3 represented by rectangles. The open rectangles represent sites that BenM binds in the absence of effectors. The shaded rectangles represent sites that BenM binds in the presence of benzoate, muconate or both effectors.
A. Benzoate → Catechol → cis,cis-Muconate → Succinyl-CoA and Acetyl-CoA

B. benPK → benM → benABCDE → catA → orf1,2 → catM → catBCIJFD

Site 1 → Site 2 → Site 3

1. BenM CatM
2. BenM primary
3. BenM CatM
4. CatM primary
replace the chromosomal \textit{catM} region using the following primers listed 5’ to 3’:

GGCTTTAACACTCTCTGGACCTTTCTTCGC (2nd Primer 1),
GTGGTGCTGGGTGCTGGGTGCTCTGATGAGTGCGCTGATATG (Primer2-His),
TATGAACCCACAAAGCCATGAA (Primer 3),
CACCACCACCACCAATATGCTGAAAAATTTACT (Primer4-his). Regions encoding the histidine tag are underlined. Two DNA fragments were generated by PCR (one with 2\textsuperscript{nd} Primer 1 and Primer2-His, the other with Primer 3 and Primer4-his). These two fragments with overlapping regions were combined and used as template for PCR amplification by 2\textsuperscript{nd} Primer 1 and Primer4-his to generate a fragment with sufficient homology to introduce the novel \textit{catM} allele (\textit{catM5550}) into the chromosome by previously described allelic replacement methods (10). The novel allele was used to replace that of a \textit{catM}-disrupted mutant, ACN541 (14), such that replacement could be detected by the loss of appropriate drug resistance.

Plasmid pBAC641, encoding CatM-EBD, was made similarly to pBAC435 encoding BenM-EBD (9). Amplification with PCR primers, 5’-

\textbf{CATATACATATGGCAA AACGGATTGCAACG-3’ (catM-IBD-Nde) and 5’-TCAATTCCTCGAGTTCGATGAGTGCGCTGATATG-3’ (CatM-3-Xho), generated a DNA fragment that was ligated to vector pET21b (Novagen) to form pBAC641. The engineered \textit{Nde}I and \textit{Xho}I restriction sites for cloning are underlined. The encoded protein, which lacks the 80 amino acids at the N-terminus of full-length CatM, has residues for a purification tag at the C-terminus (SILEHHHHHHH). PCR amplification also led to the substitution of proline for leucine at position 174. The tag at the C-terminus of BenM is LEHHHHHHH.
Expression of BenM and CatM-EBD His-tag proteins from E.coli

The effector binding domains of BenM and CatM were purified from pET-21b based plasmids, pBAC435 and pBAC641 respectively using BL21 (DE3) cells. For expression, the plasmids were transformed into BL21 (DE3) E.coli cells and plated on LB plates with 150 µg/ml of ampicillin. A colony from the plate was inoculated in 5 ml of LB and incubated with agitation for about 4 hrs at 37 °C. The entire culture was inoculated into 1.0 l of LB. Isopropyl-β-D-thiogalactosidase (IPTG) was added to a final concentration of 1.0 mM after the cultures reached an OD_{600} of about 0.5 which took about 3-5 hrs. Cultures were incubated for 12 – 16 hours and harvested by centrifugation at 6,000 x g for 10 min at 4 °C and cell pellets were stored at – 70 °C.

Purification of His-tag proteins from E.coli

Frozen cell pellets expressing CatM-EBD-His or BenM-EBD-His were suspended in 20 ml of binding buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 10% glycerol and 5 mM imidazole) at 4°C. Cells were lysed via two passes through a chilled (4 °C) French pressure cell at 15,000 psi. The cell lysate was centrifuged at 15,000 g for 15 min at 4°C, sometimes the supernatant was re-centrifuged under same condition to get rid all remaining traces of cell debris.

All purifications were done at room temperature using a High Performance liquid Chromatography (HPLC) system for the CatM-EBD-His or a Fast Performance Liquid Chromatography (FPLC) system for BenM-EBD-His. Cell lysates were applied to a 5 ml Hi-TRAP metal-chelating column that had been charged with Ni^{2+} (2.5ml of 0.1 M NiSO_{4}) and equilibrated with 5 column volumes of binding buffer.
Both proteins were eluted with a linear gradient of elution buffer at a rate of 2 ml/min. However the elution buffers used for both proteins were different, the elution buffer for the BenM-EBD-His protein constituted 20 mM Tris-HCL pH 7.9, 500 mM NaCl, 10 % glycerol and 250 mM imidazole while that of CatM-EBD-His was the same but with a spike in imidazole concentration to 500 mM. Purified protein fractions were pooled and dialyzed (twice) for 4hrs in 1.0 l of dialysis buffer using 10,000 Da molecular weight cut-off snakeskin dialysis tubing (Pierce). The dialysis buffer for BenM-EBD constituted of 20 mM Tris-HCL pH 7.9, 500 mM NaCl and 10% glycerol, while that of CatM-EBD constituted of 20 mM Tris-HCL pH 7.9, 500 mM NaCl 10 % glycerol and 250 mM imidazole. The addition of imidazole in the CatM-EBD-His dialysis buffer seemed to keep the protein in solution. Proteins were concentrated using an ultrafree S-10 centrifuge concentrator (Millipore) and stored at 4 °C until crystallization trials were set up.

**Crystallization, data collection and structure determination**

Crystals of BenM-EBD and CatM-EBD were grown using the microbatch under oil method at 15°C from the HR I-20 solution as previously described (6). Before cryofreezing in nitrogen or Freon, the crystals were soaked in solutions containing cryoprotectant and benzoate or muconate. Prefrozen crystals were transported to synchrotron beamlines in cryo shippers and transferred to the cryostream operating at 100K. To prepare a benzoate soaked BenM-EBD crystal, a large crystal was soaked in 120 mM sodium benzoate pH 4.6, 35% PEG 4000 prepared by titrating benzoate to pH 4.6 and dissolving solid PEG 4000 in a 120 mM solution of the benzoate. To prepare the
muconate derivative of BenM-EBD, a large crystal was soaked overnight in saturated
muconate in 35% PEG 4000, 10 mM Tris pH 8.0, 0.5 M NaCl. Upon standing, the
crystals soaked in the effectors (either muconate or benzoate) cracked into smaller pieces.
The largest visibly solid remaining pieces, about 0.1 mm in all dimensions, were used for
data collection. CatM-EBD crystals did not crack upon soaking of a solution of muconate
(112.5 mM in 35% PEG4000 27.5 mM NH₄SO₄, 13.75 mM Sodium acetate, 2.75 mM
Tris, 1.4 % glycerol and 34.2 mM imidazole). Data sets containing ligands were collected
at the SBC-CAT BM-19 beamline using 0.5-1° oscillation widths and 5-30 s exposures
per frame. Data were processed and scaled with the beamline version of HKL2000 (26).

For the structure determination of BenM-EBD, a selenomethionyl derivative of
BenM-EBD was prepared by transforming the BenM-EBD plasmid into the 854(DE3)
methionine auxotroph strain and growing the cells on a defined medium containing
selenomethionine (32). No other attempts were made to protect the samples from
oxidation. Mass spectrometry data (not shown) indicated that the sulfur atoms of the
methionines were well substituted with selenium. An X-ray fluorescence scan of the
crystals around the Se edge was performed at the SER-CAT beamline to identify the
appropriate wavelengths for MAD data collection. Three data sets (peak, edge and
remote, only the statistics for the peak are shown in Table 3.1) were collected and
processed with HKL2000 at the beamline. Though data sets for three different
wavelengths were collected and processed with the intention of performing MAD
phasing, only a single-wavelength resulted in an interpretable map using the
SOLVE/RESOLVE suite (www.solve.lanl.gov) with the SAD phasing approach (36, 37).
Attempts to phase using the MAD method did not result in interpretable maps. Nine
selenium sites (out of 10 final refined methionines) were identified by SOLVE (FOM 0.54) and the phases input into RESOLVE (35, 37). After one round of RESOLVE, two very incomplete chains were visible by molecular modeling. RESOLVE did not find the non-crystallographic two-fold. The appropriate rotation and translation terms for two-fold symmetry averaging were derived by interpreting the model in the program O and input into RESOLVE. The program then automatically modeled 63% of the side chains and 73% of the main chain residues using the BenM-EBD sequence at 2 Å resolution. The remaining residues were manually fitted using the program, O (19) and COOT (13). Each refinement cycle consisted of a round of model building, ARP GuiSIDE chain refinement with waters removed, appropriate waters returned, REFMAC refinement (24) and ARP water identification (29, 30). Group B factors were refined with TLS refinement with the TLS domains defined using the TLSMD web server (27). Significant drops in $R_{\text{values}}$ and $R_{\text{free}}$ occurred as a result of using the TLS refinement. Hydrogen atoms were included in the final cycles of refinement. MolProbity (12, 22) was used to evaluate proper orientations of Asn, Gln and His residues and Procheck (20) and subroutines in COOT (13) were used to identify structural issues that were corrected.

The selenomethionyl coordinates were used as a molecular replacement model for the remaining structures of benM-EBD and also for CatM-EBD using the program MolRep (40) in the CCP4 suite (1). Two copies of the monomer (either monomer A or B) were used as independent search models in the BenM-EBD structures, while only one copy was necessary for CatM-EBD. Rigid body refinement in REFMAC with the subdomain I and II refined as separate bodies, multiple rounds of refinement, model
Table 3.1. Crystallographic Data Collection and Refinement Statistics

<table>
<thead>
<tr>
<th></th>
<th>BenM-EBD</th>
<th>BenM-EBD</th>
<th>BenM-EBD</th>
<th>BenM-EBD</th>
<th>CatM-EBD</th>
<th>CatM-EBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeMet peak</td>
<td>Native</td>
<td>EBD+CCM</td>
<td>EBD+BEN</td>
<td>EBD</td>
<td>EBD+CCM</td>
<td></td>
</tr>
<tr>
<td>Beamline¹,</td>
<td>22-ID,</td>
<td>19-BM,</td>
<td>19-BM,</td>
<td>19-BM,</td>
<td>19-BM,</td>
<td>19-BM,</td>
</tr>
<tr>
<td>wavelength (Å)</td>
<td>0.97935</td>
<td>0.97937</td>
<td>1.00727</td>
<td>1.00727</td>
<td>1.00727</td>
<td>0.97934</td>
</tr>
<tr>
<td>Cell constants</td>
<td>a=65.1,</td>
<td>a=64.4,</td>
<td>a=64.7,</td>
<td>a=64.4,</td>
<td>a=66.2,</td>
<td>a=65.6,</td>
</tr>
<tr>
<td>(Å), Space Group</td>
<td>b=66.7,</td>
<td>b=66.6,</td>
<td>b=65.8,</td>
<td>b=69.5,</td>
<td>b=75.5,</td>
<td>b=73.1,</td>
</tr>
<tr>
<td></td>
<td>c=117.8,</td>
<td>c=116.5,</td>
<td>c=118.0,</td>
<td>c=116.9,</td>
<td>c=102.2,</td>
<td>c=102.6,</td>
</tr>
<tr>
<td></td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>I222</td>
<td>I222</td>
</tr>
<tr>
<td>Resolution range</td>
<td>58.93-2.00</td>
<td>46.32-1.91</td>
<td>30.0-1.8</td>
<td>59.76-2.05</td>
<td>60.75-1.90</td>
<td>48.8-2.16</td>
</tr>
<tr>
<td>(Å)</td>
<td>(2.05-2.00)</td>
<td>(1.97-1.91)</td>
<td>(1.81-1.95)</td>
<td>(2.05-2.10)</td>
<td>(1.95-1.90)</td>
<td>(2.21-2.16)</td>
</tr>
<tr>
<td>Completeness of</td>
<td>99.5 (97.6)</td>
<td>73.56 (18.9)</td>
<td>88.5³(93.9)</td>
<td>95.7 (97.5)</td>
<td>94.4 (97.5)</td>
<td>98.0 (87.04)</td>
</tr>
<tr>
<td>data² (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rmerge⁴ (%)</td>
<td>6.9 (22.1)</td>
<td>3.3 (43.4)</td>
<td>5.3 (53.8)</td>
<td>6.8 (59.9)</td>
<td>5.6 (73.0)</td>
<td>7.9 (53.7)</td>
</tr>
<tr>
<td>Average I/σI</td>
<td>49.6 (9.6)</td>
<td>31.2 (1.4)</td>
<td>25.8 (3.9)</td>
<td>25.2 (3.8)</td>
<td>29.7 (7.5)</td>
<td>42.6 (4.1)</td>
</tr>
<tr>
<td>Refined</td>
<td>0.80-1.22</td>
<td>0.32-0.59</td>
<td>0.22-0.31</td>
<td>0.50-0.58</td>
<td>0.13-0.31</td>
<td>2.16-2.67</td>
</tr>
<tr>
<td>mosaicity⁵</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rvalue⁶ (%)</td>
<td>18.1 (18.7)</td>
<td>19.0 (30.1)</td>
<td>20.7 (34.2)</td>
<td>18.0 (24.8)</td>
<td>17.2 (20.9)</td>
<td>18.2 (22.7)</td>
</tr>
<tr>
<td>Rfree⁷ (%)</td>
<td>22.7 (23.4)</td>
<td>23.4 (32.0)</td>
<td>25.0 (40.7)</td>
<td>26.4 (28.2)</td>
<td>24.0 (27.8)</td>
<td>24.2 (29.7)</td>
</tr>
</tbody>
</table>

¹ Data sets were collected at the Advanced Photon Source, Argonne, IL. Beamline 19-BM was operated by Structural Biology Consortium Collaboratory Access Team (SBC-CAT) and 22-ID was operated by the SouthEast Regional Collaboratory Access Team (SER-CAT).
² Numbers in parentheses indicate the values for the highest resolution bin.
³ A large swath of data, approximately 25% of the shell, is missing between 2.46-2.15 Å resolution due to the presence of a strong ice ring in that range that resulted in the rejection of many reflections.
⁴ Rmerge is the unweighted R value on I between symmetry related reflections.
⁵ The mosaicity is that reported from Scalepack (26). In some cases, the mosaicity was refined in batches, and thus the range (minimum and maximum) of values is reported.
⁶ Rwork = Σₜₕₗ [Fobs(hkl)-Fcalc(hkl)]/ Σₜₕₗ Fobs(hkl) for reflections in the working data set.
⁷ Rfree is the R value for the 5% of the reflections in the cross-validation data set that were not used in refinement.
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>r.m.s.d bond lengths (Å)</td>
<td>0.007</td>
<td>0.007</td>
<td>0.008</td>
<td>0.007</td>
<td>0.007</td>
<td>0.008</td>
</tr>
<tr>
<td>r.m.s.d bond angles (deg.)</td>
<td>1.05</td>
<td>1.23</td>
<td>1.09</td>
<td>1.12</td>
<td>1.08</td>
<td>1.16</td>
</tr>
<tr>
<td>Mean B (Å²)</td>
<td>26.4</td>
<td>28.4</td>
<td>30.7</td>
<td>35.9</td>
<td>19.8</td>
<td>38.5</td>
</tr>
<tr>
<td>Ligands</td>
<td>10 seleno-methionine, 1 sulfate, 1 CCM, 2 benzoate, 5 benzoate</td>
<td>4 sulfate, 1 CCM, 1 Cl⁻</td>
<td>1 sulfate, 1 acetate, 1 CCM, 1 sulfate</td>
<td>2 sulfate, 1 sulfate, 1 acetate</td>
<td>1 sodium</td>
<td></td>
</tr>
<tr>
<td>Water molecules</td>
<td>487</td>
<td>579</td>
<td>481</td>
<td>588</td>
<td>359</td>
<td>218</td>
</tr>
<tr>
<td>Phi/Psi distribution</td>
<td>92.4/7.0/0.5</td>
<td>92.9/6.8/0.0/0.3</td>
<td>90.8/9.2/0.0/0.0</td>
<td>91.4/8.6/0.0/0.0</td>
<td>95.2/4.8/0.0/0.0</td>
<td>95.2/4.8/0.0/0.0</td>
</tr>
</tbody>
</table>
building and water identification as described previously for the selenomethionyl
derivative were performed on all of the data sets. CatM-EBD required additional
modeling to compensate for differences in the amino acids sequences between CatM and
BenM. The inclusion of the effectors muconate and benzoate in REFMAC refinement
necessitated the creation of geometry files for proper refinement. The occupancies of the
ligands or side chains were not refined. Final refinement statistics are reported in Table
3.1. Buried surface areas were calculated using the program GetArea1.1 via a world wide
web (15). Figures were drawn using DeepView/Swiss-PDB Viewer v3.7 (17) and
rendered with PovRay v3.6 (www.povray.org). Structural validations were performed
with Procheck (20) and MolProbity (12).

RESULTS

Structural characterization of BenM-EBD, CatM-EBD and their ligand complexes

The EBDs of CatM and BenM lack the 80 N-terminal residues encompassing the
DNA binding domain (6). A six-histidine tag at the C-terminus facilitated purification.
This tag does not interfere with the ability of BenM to function in vivo or to bind
effectors in vitro (4, 9). To confirm that the purification tag does not interfere with CatM
function, an allele encoding the modified protein (catM5550) was introduced into the A.
baylyi chromosome. The resulting mutant grew at wild-type rates with benzoate as the
sole carbon source (data not shown). Thus, the histidine-tagged CatM, like its BenM
counterpart, regulates gene expression during growth on benzoate.

The truncated EBD proteins were maintained in high concentrations of NaCl (0.5
M) and glycerol (10% v/v). For CatM-EBD, imidazole (circa 250 mM) was also required
to maintain solubility. These conditions required crystallization by batch mode under oil rather than by vapor diffusion methods. Crystals of BenM-EBD, grown as reported previously (6), had two monomers per asymmetric unit (Table 3.1). Crystals of CatM-EBD, grown with the same solution (Hampton Research Crystal Screen I, condition 20), formed in a different space group with one monomer per asymmetric unit (Table 3.1). The SOLVE/RESOLVE suite (37) was used to determine the structure of BenM-EBD with a single-wavelength data set from a selenomethionyl-modified version of the protein. Despite collecting multiple wavelength data sets with the intention of doing a MAD structure determination (18), only the first data set collected at the selenium peak wavelength resulted in a structure. Crystal decay during data collection may have reduced the quality of the later data sets.

Initial attempts to prepare ligand complexes of BenM-EBD were unsuccessful. Crystals were soaked with effectors at concentrations at least 10-fold higher than the $K_d$, previously shown to be 0.12 mM for muconate and 1 mM for benzoate (9). Although diffraction data were collected from crystals that did not crack during the soaking procedure, the difference Fourier maps failed to reveal density features indicative of bound effectors. Nevertheless, several sulfate and chloride ions were evident in what was later identified as the primary effector-binding site (data not shown). The high level of negatively charged ions such as chloride, sulfate and acetate in the crystallization solutions necessitated the use of high effector concentrations that might help displace solvent anions bound to the protein. The effectors were introduced slowly to crystals of varying size, ranging from small to as large as approximately 0.3 x 0.3 x 1 mm.
These procedures caused the BenM-EBD crystals to crack, and the effectors altered the unit cell lengths in the crystals by as much as 3 Å (Table 3.1). Nevertheless, shards of the broken crystals used in X-ray diffraction studies yielded structures of the effector-bound BenM-EBD complexes. In some cases the shards diffracted to a higher resolution than the native crystals. Improved resolution may result from more efficient heat transfer during freezing of the smaller crystals. This possibility is supported by the observation that the refined crystal mosaicity values for the ligand complexes were lower (~0.3°) than for the native protein (Table 3.1).

The CatM-EBD crystals did not crack. However, the CatM-EBD soaked with muconate had a refined mosaicity (>1°) substantially higher than in the absence of effectors. The overall conformational changes between the proteins associated with different crystal data sets were relatively small and mainly involved interdomain movements. Thus, molecular replacement was performed for BenM and CatM using the BenM-EBD selenomethionyl-derived structure as the search model. The two domains of the EBD, discussed below, were treated as rigid bodies. All structures were refined to acceptable crystallographic standards. The r.m.s. deviations were less than 0.009 Å and 1.3° for bond distances and angles, respectively. Statistics associated with each X-ray structure are shown in Table 3.1. Two molecules of BenM-EBD in the crystallographic asymmetric unit are related by a non-crystallographic two-fold axis of symmetry. In contrast, CatM-EBD structures contain a single monomer in the asymmetric unit.
Protein fold of CatM-EBD and BenM-EBD

As shown in Figure 3.2A and B, the BenM-EBD and CatM-EBD monomers consist of an α/β structure with two Rossmann-like folds. Each structure has 9 β-strands and 9 helices, four of which are $3_{10}$-helices. These two EBDs are nearly identical in structure with an overall backbone r.m.s. deviation of 1.57 Å. As noted for the regulatory domains of previously characterized LTTRs, these structures are essentially those of periplasmic binding proteins (31). Each EBD has two domains separated by an inter-domain hinge comprised of the central regions of two antiparallel β-strands (β4 and β9). Domain I includes the N-terminal EBD region, amino acids 87 to 161 of the full-length protein, and amino acids 268 to the C-terminus. Residues 81 to 87 in these truncated regulatory proteins are presumably part of a linker helix that connects the DNA-binding domain to the EBD in the full-length proteins (23). Domain II contains the intervening sequence between the two domain I regions, amino acids 162 through 267. Data from different crystals suggest that domains I and II can move relative to each other by rotation around the pivotal hinge-like β-strands. The electron densities of the N- and C-terminal ends of the EBD proteins vary between different crystal data sets, which limits the accuracy of modeling these terminal residues.

Central muconate-binding site in BenM-EBD and CatM-EBD

A muconate-binding site was identified between domains I and II of BenM-EBD and CatM-EBD. The density for a muconate molecule bound to a CatM-EBD monomer was evident from difference Fourier maps of data sets from crystals prepared with and
FIG 3.2 Ribbon representation of BenM-EBD and CatM-EBD Panel A illustrates the α-carbon backbone of BenM-EBD monomer B of the EBD without bound ligands in a ribbon representation, colored blue to red, amino terminus to carboxy terminus. The 9 helices and 9 β -strands are denoted H1-H9 and β1- β9. Helices H1, H4, H5 and H8 are 3₁₀-helices that are variably structured in the different atomic structures depending on which monomer is in the asymmetric unit or the ligand state. The monomeric unit can be divided into two domains, domain I includes the N-terminal half of the protein, H1-αH3 and strands β1- β4, and the C-terminus, β9 and αH9. Domain II includes elements β4 through β9. Panel B illustrates the α-carbon backbone of the CatM-EBD monomer. Panel C illustrates the dimeric structure of BenM-EBD with bound benzoate and muconate. The effectors muconate and benzoate are shown in space filling mode. The bound muconate molecule is centrally located in BenM-EBD while the bound benzoate molecules are internally within the β-sheet structure. The dimer is the asymmetric unit in the BenM-EBD P2₁2₁2₁ crystal form. Benzoate was not in the crystallization solutions or the artificial mother liquors/cryoprotectant and is presumed to have originated as contaminant in the purification process. Panel D illustrates the dimeric structure of CatM-EBD with bound muconate. The muconate is shown in space filling mode. The dimeric structure was generated by applying a crystallographic symmetry transformation to derive the second molecule of the dimer. The bound muconate molecule is also centrally located within CatM-EBD. Panel E illustrates the monomeric structure of chain B of BenM-EBD emphasizing the primary and secondary binding sites with bound muconate and benzoate respectively.
without muconate. The location of this muconate-binding site is shown for a CatM-EBD dimer (Figure 3.2C). For BenM-EBD, waters that were initially modeled into the structure obscured the ligand-binding sites. As refinement progressed, the water models were omitted and the muconate molecule could be readily modeled using the difference density maps. The position of the muconate-binding site in BenM-EBD, designated the primary effector-binding site, was comparable to that in CatM-EBD.

In the crystal forms of BenM-EBD, the two monomers in the crystallographic asymmetric unit are slightly different from one another. This difference was notable in one crystal soaked with high concentrations of muconate in which the r.m.s. deviation between the positions of the backbones averaged to 1.25 Å over the entire length of the monomers. As shown in Figure 3.2D, muconate was bound to the primary binding site in only one of the two monomers in the crystal. Surprisingly, a significant density feature in the hydrophobic core of domain II could not be adequately modeled with waters. Further analysis and additional studies of benzoate-soaked crystals of BenM-EBD, described next, suggest that benzoate was bound to a secondary binding site distinct from the location of the muconate, Figure 3.2D and 2E. The benzoate in the muconate-soaked complex most likely came from the muconate sample itself, as this compound was derived enzymatically from benzoate. Alternatively, benzoate might be a minor contaminant of the imidazole used during protein purification.

**Secondary benzoate-binding site in BenM-EBD but not CatM-EBD**

Difference Fourier maps were computed for BenM-EBD structures prepared with benzoate. These analyses revealed benzoate bound to the secondary effector-binding site
in BenM-EBD (Figure 3.2D and 2E) and also to the primary effector-binding site, previously identified as the site for binding muconate. All density patterns were consistent with this interpretation, and the refined structure resulted in local effector environments that only had appropriate van der Waal’s contacts. Comparable attempts to detect benzoate-bound complexes of CatM-EBD were unsuccessful (data not shown).

The key residues involved in protein-effector interactions in the primary effector-binding site are nearly identical between BenM and CatM, whereas those in the secondary site are not. As discussed below, structural analysis predicts that CatM-EBD binds benzoate in the primary but not the secondary effector-binding site.

**Protein-Effector interactions in the primary binding site of BenM-EBD and CatM-EBD**

Specific interactions between muconate, in the primary effector-binding site, and residues of BenM-EBD or CatM-EBD are remarkably similar, Figure 3.3A and B. The numbering of amino acids corresponds identically between the BenM- and CatM-EBD structures. While most ligand-protein contacts are direct, there are some bridged interactions involving water. Designations for these water molecules derive from the number of the nearest amino acid. As shown in Figure 3.3A and B, one carboxylate atom (O1) of muconate hydrogen bonds to the side chain oxygen of Ser99 and water 195, which in turn forms a hydrogen bond with Tyr195. Another carboxylate atom (O2) of muconate forms hydrogen bonds to the side chain oxygen and the amide nitrogen atoms of Ser99. Water 227 bridges the O2 oxygen of muconate to the amide nitrogen of
FIG. 3.3. Stereo views of the primary effector binding site and the secondary putative activation site. Panel A shows neighboring atoms around the BenM-EBD muconate binding site and the corresponding difference electron density contoured at 3.0 $\sigma$ (structure BenM-EBD+CCM). Panel B shows the same region of CatM-EBD with muconate bound with its corresponding difference electron density contoured at 3.0 $\sigma$ (structure CatM-EBD+CCM). The muconate difference density was much more pronounced in the CatM-EBD+CCM than the BenM-EBD+CCM structure. The CatM-EBD+CCM difference density was calculated before muconate was ever introduced into the refinement of the model. Panel C shows the neighboring atoms and electron density around the benzoate bound in the primary effector binding site of BenM-EBD+BEN. This density is much less clearly defined than that for the muconate density of BenM-EBD+CCM. Panel D shows the difference electron density (3.0 $\sigma$) and residues around the secondary benzoate binding site. Glu 162 and Arg 160 are included in Panel D to show the close relationship between the primary and secondary sites. Difference electron densities for the BenM-EBD structures were calculated by removing the effector atoms from the coordinate set, performing 25 cycles of refinement in REFMAC, then calculating a difference Fourier map.
Val227. At the other end of the muconate molecule, the O4 of the second carboxylate group interacts with the amide nitrogen and the side chain oxygen of Thr128 and water 201. Water 201 further interacts with the main chain oxygen of Pro201, and the main chain amide nitrogens of Thr128 and Lys129. The O3 carboxylate oxygen atom interacts with the main chain nitrogen of Phe203 and two water molecules, waters 198, which hydrogen bonds to Ser198, and water 201.

Hydrophobic interactions occur between the aliphatic carbon atoms of the muconate and the phenyl ring of Phe203, and the aliphatic side chains of Val97 and Arg146. CatM differs from BenM in this region at only one amino acid, having a Ser rather than Gly at position 98. The hydroxyl side chain of Ser 98 in CatM, which does not interact directly with muconate, appears to help stabilize water 99. In BenM, a water molecule resides at a position corresponding to the Ser98 side chain of CatM.

In contrast to these specific interactions between all parts of muconate and the primary effector-binding site of the protein, benzoate fits less snugly within this binding pocket. As shown for BenM-EBD in Figure 3.3C, benzoate interactions with the protein consist of a subset of those used to bind muconate. The carboxyl group of the benzoate interacts with the hydroxyl side chain of Ser99. Hydrophobic interactions of the benzyl ring are with Phe203, Leu147, and Arg146. Water 99 acts as a bridge to the amide nitrogen of Ser99, and water 203 bridges to the amide nitrogen of Phe203. Many more water molecules are in the vicinity of benzoate than muconate, and the ligand fit is relatively loose for benzoate. Additional data from crystals in different space groups, soaked with varying concentrations of benzoate, show similar hydration patterns and protein-effector interactions (data not shown).
**Benzoate-BenM-EBD interactions in the secondary effector-binding site**

In the secondary effector-binding site of BenM-EBD, the carboxyl group of benzoate forms salt bridges with the side chain of Arg160 and hydrogen bonds to the side chain hydroxyl of Tyr293 and the main chain amide nitrogen of Leu104. The carboxyl group of the benzoate further hydrogen bonds to water 219. Waters 219 and 719 distantly interact with Ser246 (3.45 Å and 3.55 Å) and Glu162 (3.37 Å and 3.24 Å). Hydrophobic interactions are significant between the benzoate ring and the hydrophobic side chains of residues Leu100, Leu105, Ile108, Phe144, Leu159, Ile269, and Ile289.

**DISCUSSION**

**Conformational changes in BenM- and CatM-EBD associated with transcriptional regulation**

Overall, BenM- and CatM-EBD are similar to the four previously characterized structures of LTTR regulatory domains (5, 23, 34, 39). Two of these regulators, DntR and CbnR, belong to an LTTR subclass involved in bacterial aromatic compound degradation that are similar to BenM and CatM (38). In contrast, there is minimal sequence similarity between the regulatory regions of CysB or OxyR and the ADP1 regulators. For example, sequence identity between BenM-EBD and the comparable regions of CysB and OxyR are only 12 and 21%, respectively (Figure 3.4). Therefore, the high level of structural conservation is striking. The carbon backbone of a BenM-EBD monomer aligned with the comparable backbones of DntR, OxyR and CysB, exhibits an average r.m.s. deviation of 1.51, 6.07 and 7.43 Å, respectively. Many biochemical and genetic studies led previously to the hypothesis that the effector-binding site of LTTRs is
FIG. 3.4. shows a multiple sequence alignment and secondary structure assignment. Helical regions of the secondary structure of BenM-EBD are shown as cylinders and β-strands are shown as arrows. 3₁₀-helices (H1, H4, H5, and H8) are drawn as three balls on a string. The coloring scheme for secondary structure elements is blue to red, amino terminal end to carboxy terminal end. The amino acid sequences used for the sequence alignment are as follows: (1) BenM (ADP1; O68014), (2) CatM (ADP1; PO7774), (3) CatR (*Pseudomonas putida*; P20667) (4) HcaR (*Escherichia coli*; Q47141), (5) AlsR (*Bacillus subtilis*; Q04778), (6) CbeR (*Burkholderia* sp. NK8; Q9AQS6), (7) CbbR (*Rhodobacter sphaeroides*; P52690) (8) ClcR (*Pseudomonas putida*; Q05840), (9) TcbR *Pseudomonas* sp. strain P51; P27102), (10) TfdS (*Ralstonia eutropha* JMP134; Q46M54), (11) CbnR (*Ralstonia eutropha*; Q9WX7), (12) DntR (*Burkholderia* sp. DNT; Q7WT50), (13) OxyR (*Escherichia coli*; P0ACQ4), (14) CysB (*Salmonella typhimurium*; P06614). The alignment of DntR, OxyR, and CysB sequences was performed by structural superposition of the atomic coordinates with the BenM-EBD structure using Swiss-PDB viewer with manual addition of the aligned sequences to the alignment of the homologues. Absolutely conserved residues among the BenM homologues have black backgrounds and white text; highly conserved residues have grey backgrounds. Residues denoted with a ✗ symbol are within contact distance of the *cis,cis*-muconic acid in the primary effector binding site. The Tyr and Arg 160 that form hydrogen bonds to the benzoate are highlighted purple. Residues denoted with a ● symbol are within contact distance of the benzoate molecule in the secondary site. Conserved Glu 162 and Ser 246, discussed in the text, are denoted by red highlighted letters.
located at the interface between the two regulatory domains of the monomer (2, 11, 21, 31, 33, 34, 39). These studies of BenM-EBD and CatM-EBD are the first to confirm the structure of any LTTR in complex with its cognate effector. As predicted, a primary effector-binding site was located between domain I and II (Figure 3.2). When muconate binds to this site, it allows BenM and CatM to activate transcription of the ben and cat genes.

Muconate binding causes the two EBD domains to move towards each other, connected by two hinge-like anti-parallel β–strands (Figure 3.2). This movement is affected by helical dipole moments. Ser99, Thr128, Phe203, and the water-bridged Val 227 are at the amino terminal ends of helices H1, αH3, αH6, and αH7 (Figure 3.5A). These helices are oriented such that their dipole moments interact with the negative charge of the muconate carboxylates via the amino terminal ends of the helices. This source of charge-mediated interaction with muconate is significant, since there are no nearby positively charged amino acids. The Nε of Arg146 is distant, 3.64 Å away, from the nearest carboxyl group of muconate, Figure 3.5. Moreover, Arg146 is salt-bridged to Glu162 in both BenM- and CatM-EBD structures. Any mechanism whereby the dipole moments of these helices are enhanced, whether by local mutations or ligand binding, could increase transcriptional activation. The binding of muconate, especially to CatM, also alters hydrogen-bonding patterns. For example, muconate displaces the usual location of Arg146 in CatM-EBD, thereby allowing it to interact with Ser99, Glu162 and Ser267. These added interactions among residues from both domains should help stabilize the transcriptionally active ligand-bound form of the regulator.
FIG. 3.5. Schematic representation of the effector binding sites. Panel A illustrates the primary effector binding site of BenM-EBD containing muconate with hydrogen bonds and residues in contact with the ligand shown. Dashed lines represent hydrogen bonds. The long distance of the Arg 146 Ne atom to the O1 of muconate is shown for comparison as a dash-dot line. Distances in Ångströms are between non-hydrogen atoms. Helices that are nearby are schematically shown as cylinders—colored coded as in Figures 1 and 2. β-strands have been omitted for clarity. An attempt to confer depth is represented by the thickness of the lines. Panel B illustrates the benzoate in the secondary binding site. H1 is a 3₁₀-helix associated with both the primary and secondary benzoate binding sites.
The binding of muconate to BenM-EBD in the primary effector-binding site results in significant conformational changes, with four helices moving several Ångstroms with respect to their positions in the absence of effectors. However, the changes that occur in this region when benzoate binds to this same primary site are much less pronounced. The effector-binding responses for benzoate and muconate are different in the primary site. Consistent with such differences, BenM and CatM differ in their ability to respond to benzoate as an effector despite having nearly identical primary effector-binding sites (Figure 3.3). Since CatM does not respond to benzoate as an inducer, binding of this compound to the primary effector-binding site may not be sufficient to enable transcriptional activation. BenM activation of transcription in response to benzoate as the sole effector may depend on this compound occupying both the primary and secondary effector-binding sites in BenM.

Role of the secondary effector-binding site in transcriptional synergism

The most surprising finding from our structural studies is the presence of two distinct effector-binding sites in BenM. The secondary binding site would not have been predicted from the known LTTR structural data. This site, which binds benzoate but not muconate (Figure 3.5B), most likely produces the synergistic response of BenM by accentuating the signal from muconate bound in the primary site. Synergistic activation of BenM may involve two mechanisms, displacement of the hydrophobic core and surrounding residues and/or a charge relay system. In the secondary effector-binding site, benzoate resides near helices H1 and αH2. Leu100, on H1, is well positioned to
assist muconate binding when benzoate is in the secondary site (Figure 3.5). Its neighbor, Ser99, provides two contacts to muconate and would be affected by conformational changes near Leu100. Another key residue is Glu162, which appears able to pivot from Arg160 (at the secondary site) to Arg146 (at the primary effector site). Glu162 is centrally located in BenM-EBD in the middle of β-strand 4, which spans both domains. The importance of this hinge-like region in effector response is supported by mutational analysis of comparable residues in other LTTRs such as CysB (21) and CbbR (11).

The charge relay between Arg160, Glu162, and Arg146 is likely to be important. When benzoate is bound, Arg160 forms a salt bridge with the carboxyl group of the benzoate. The superimposition of the CysB structure on BenM showed that regions implicated in inducer response in CysB (21) corresponded to regions of this secondary binding site. For instance Tyr164 in CysB would be within approximately 2.5 Å of the benzoate moiety. The hypothesis that changes in this local region can significantly affect transcriptional activation is also supported by recent mutational studies of a distantly related LTTR, OccR (2). When a model of the OccR monomer using CbnR as a template was computed, some of the residues involved in the inducer response were also in close proximity to the secondary binding site, such as Pro149, Ser123, Ala111, and Leu120. In fact, Pro149 was also in close proximity to the proposed primary binding site. Some of the OccR double mutants such as L120F/R202W acted synergistically. R202 in OccR is located in the primary binding site and structurally corresponds to residue F203/Asn202 in BenM. It is attractive to postulate that the OccR variants could be replicating the types of changes in charge distribution that are involved in the synergistic response of BenM.
Our model predicts that the form of BenM tetramer able to activate the highest level of *benA* transcription is comprised of monomers each bound to muconate in the primary effector-binding site and benzoate in the secondary site. The conformation of the tetramer bound to both effectors should be slightly different from that bound to muconate or benzoate alone. Such conformational differences would account for altered transcriptional regulation. BenM-effector interactions were previously studied by fluorescence emission spectroscopy (9). While these studies could not determine whether benzoate and muconate bind to the same or different sites, they clearly show that BenM has a higher affinity for muconate than benzoate as the sole effector.

In the primary effector-binding site, there are stronger specific interactions between the binding-pocket residues and muconate than with benzoate (Figure 3.3A compared to 3.3C). These interactions can account for the relative values of the apparent binding constants for benzoate and muconate. Furthermore, the ability of both compounds to bind to this primary site enables competitive binding. Such competition in the primary site could explain the observation that benzoate, when present at half-saturation concentration, lowers the apparent binding affinity of muconate (9).

Based on the structural environment around the secondary binding site for benzoate (Figure 3.3D), we would expect the affinity for benzoate in this site to be higher than for the primary site. We would also expect from analysis of the van der Waal’s packing in this secondary site that adequate space is available for either *ortho* or *meta* position chlorides on the benzoate ring. Based on the BenM-EBD structure, meta-substituted compounds would be expected to bind best. Consistent with this analysis, fluorescence emission spectroscopy indicated that BenM-EBD binds several benzoate
analogs, including $o$-chlorobenzoate, $m$-toluate, $m$-chlorobenzoate and isophthalate. The affinity for $m$-chlorobenzoate was approximately 3-fold greater than for benzoate (7).

**Ligand-mediated global effects on tetrameric structure**

The binding of muconate to a BenM-EBD monomer results in the movement of main chain and side chain atoms. While this movement is at most several Ångstroms, small changes can confer a large global effect on the native tetrameric structure. One feature of LTTRs, seen in the full-length CbnR structure, is that the DNA binding domains are spatially on one face of the tetramer (23). The tetramer is effectively a dimer of asymmetric dimers (Figure 3.6). Conformational changes in two monomers on one side of the tetramer are additive. By the two-fold symmetry relationship, the two DNA binding domains are drawn together in response to effectors. Rearrangement of the monomer-monomer and dimer-dimer interfaces may enhance the conformational changes that occur. This model accounts for the effector-mediated shift in DNA binding/bending observed for many LTTRs. For example, in the benA promoter region, either BenM or CatM binds two recognition sequences in the absence of effectors (Figure 3.1B, sites 1 and 3). In the presence of effector(s), binding shifts to two sequences that are closer together (Figure 3.1B, sites 1 and 2) (4, 14).

Although the regulators function as tetramers, the EBD monomers form dimers in solution (6). The interface of a BenM-EBD dimer can be studied directly since each crystallographic asymmetric unit contains two slightly different monomers. The two monomers interact through helices $\alpha$H2, $\alpha$H7, strands $\beta$2 and $\beta$6 and a random coil region, residues 249-254. By applying appropriate crystallographic symmetry
transformations, a CatM-EBD dimer can be generated that uses the same interfaces as BenM-EBD. The buried surface area of the contact, calculated for the muconate complex of BenM-EBD, is 1306 Å². In comparison, DntR, which has a similar monomeric structure to BenM, has extensions of two β-strands in domain II, the equivalent of strands β2 and β6 in BenM. Relative to BenM-EBD they create an extended pleated sheet that spans two monomers and contribute to increased buried surface contact area between the DntR monomers, 1539 Å² (34). Furthermore, the regulatory domains of DntR are rotated relative to those of BenM-EBD (or CysB and OxyR), which should generate additional differences in the quaternary structures of BenM and DntR. From an evolutionary standpoint, it seems reasonable for the interfaces of distinct LTTRs to differ substantially to avoid metabolically inactive mixed dimers.

**Are dual effector-binding sites unique to BenM?**

Currently, BenM is the only transcriptional regulator known to have a synergistic response to different effectors. However, no studies have addressed the mechanisms by which other LTTRs respond to multiple effectors. Many regulators such as CbeR and CbnR, which are involved in the degradation of chlorobenzoate and related aromatic compounds, recognize more than one inducer (16, 25). In CbeR, the region corresponding to the secondary effector-binding site of BenM has residues capable of interacting with the carboxyl group of benzoate, such as Arg160 and Tyr293, as well as significant conservation of the hydrophobic residues surrounding the binding pocket (Figure 3.4). In contrast, this region of CbnR has no voids and is filled with hydrophobic side chains. A secondary effector-binding site, in some but not all BenM homologs, is likely to contribute to transcriptional synergism in additional regulatory systems.
FIG. 3.6. Schematic model for the conformational changes in the full-length BenM or CatM tetramer in response to binding of its effector cis,cis-muconate. Each monomer is colored pink and yellow for domains I and II respectively. The light blue and dark blue cylinders represent the coiled-coil linker and DNA binding domains respectively. The 2-fold symmetry axis of the tetramer is represented by a black circle. The symbol \( \times \) represents the effector. The structures of BenM-EBD and CatM-EBD with and without bound muconate bound show that the domains contract toward one-another upon effector binding in directions shown by the short black arrows. The conformational changes within each monomer are propagated throughout the tetramer to cumulatively move the DNA binding domains closer toward the center of the tetramer represented by the long black arrows. Binding of effector thus shifts BenM's recognition of the \( benA \) promoter from the extended conformation that recognizes sites 1 and 3 of the DNA to the more compact structure that would more favorably recognize sites 1 and 2 of the promoter.
Synergism provides a rapid method for the cellular integration of multiple metabolic signals (4, 9). These comparative structural studies of BenM-EBD and CatM-EBD suggest that changes in charge distribution result from the binding of a second ligand. Dual-effector binding appears to promote conformational change in LTTR structure that is similar to amino acid substitutions in a wide-variety of mutants selected by their ability to increase transcriptional activation.
ACKNOWLEDGMENTS

The data set for the selenomethionyl BenM-EBD was collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source (APS), Argonne National Laboratory with the assistance of John Rose and Zhi-Jie Liu. Data sets for the BenM-EBD and CatM-EBD and effector complexes were collected at the Structural Biology Center Collaborative Access Team (SBC-CAT) 19-BM beamline at the APS with the assistance of Santiago Lima and Michelle Momany. The assistance of the staff at both beamlines is greatly appreciated. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. This work was supported by grants from the National Science Foundation, MCB-0346422 (to CM) and MCB-0516914 (to ELN).
REFERENCES


8. **Clark, T. J., C. Momany, and E. L. Neidle.** 2002. The *benPK* operon, proposed to play a role in transport, is part of a regulon for benzoate catabolism in *Acinetobacter* sp. strain ADP1. *Microbiology.* **148:**1213-23.


CHAPTER 4

STRUCTURAL BASIS FOR THE ALTERED REGULATORY ROLES OF THREE LYSR-TYPE TRANSCRIPTIONAL REGULATOR VARIANTS OF BENM AND CATM: BENM(R156H/T157S), CATM(V158M) AND CATM(R156H)1

ABSTRACT

BenM and CatM are LysR-type transcriptional regulators that control benzoate degradation by the bacterium, *Acinetobacter baylyi* ADP1. Both proteins play overlapping roles in the regulation of multiple genes. Recent studies have identified particular amino acid substitutions in these proteins that affect their regulatory function. Atomic structures of the effector binding domains of three variants, BenM(R156H/T157S), CatM(V158M) and CatM(R156H) were determined at resolutions between 1.8 Å and 2.5 Å. BenM(R156H/T157S) and CatM(R156H) increase effector-independent expression of *benA* and *catB* respectively compared to their wild-type counterparts. CatM(V158M) and CatM(R156H) increase cis,cis-muconate-dependent transcription of *benA* in relation to the wild-type CatM. The crystal structures of the effector binding domain of these variants were compared to the wild-type structures in light of their different properties in altering gene regulation. Here, we report the role that specific residues of BenM and CatM play in affecting transcriptional regulation. These residues, R156 for BenM and CatM and are not located directly within the effector binding site. However, through interactions with residues that contribute to binding in the effector binding pocket and an important helix, they could serve as anchor points of the protein in inducing conformational changes that alter the regulatory roles of these proteins. The substitutions R156H of BenM and CatM alter the local hydrogen bonding patterns and create conformational shifts that mimic the activated form of the wild-type protein bound to its cognate effector, cis,cis-muconate. The structures of these variants also suggest that these amino acid substitutions could alter transcriptional regulation by changing the orientation of residues that constitute the tetrameric interface of the protein.
INTRODUCTION

Variants of transcriptional regulators that have altered responses are powerful tools for the study of gene regulation. BenM and CatM belong to the large and diverse family of LysR-type transcriptional regulatory proteins. Both proteins control the expression of multiple genes needed for degradation of benzoate in the soil bacterium *Acinetobacter baylyi* ADP1 (6, 8, 12, 20). BenM and CatM are 59% identical and activate transcription of target genes in response to a metabolite of benzoate degradation, cis,cis-muconate (hereafter referred to as muconate). BenM, unlike CatM, also responds to benzoate. BenM, the primary regulator of the *benABCDE* operon is required for benzoate consumption (Figure 4.1) (8). Strains that lack a functional copy of *benM* cannot grow on benzoate (12).

Two variants of CatM, CatM(V158M) and CatM(R156H) allow increased CatM-activation of the *benABCDE* promoter in response to muconate relative to the wild-type CatM (12). These variants had different effects on the *benA* and *catB* promoters. CatM(V158M) increased muconate-dependent transcription of *benA* in relation to wild type CatM but decreased that of *catB*. On the other hand, CatM(R156H) activates high *cat* operon expression in the absence of muconate (12, 20). Similar to the R156H substitution in CatM, a homologous substitution (R156H) was generated and characterized in BenM (9). This mutation allowed increased activation of the *ben* genes in the absence of effectors compared to the wild-type BenM. The double mutant of BenM(R156H), BenM(R156H/T157S) showed similar properties to BenM(R156H) in activating the *ben* genes in the absence of effectors (9). These mutations R156H, T157S and
FIG. 4.1. Benzoate degradation pathway in ADP1 is regulated by CatM and BenM and the effectors benzoate and muconate (A). The chromosomal ben and cat genes encode proteins for benzoate catabolism (B). BenM and CatM regulate transcription in the regions shown (black circles). In A. baylyi strains lacking benM, growth on benzoate is not possible, however variants CatM(R156H) and CatM(V158M) can activate the expression of the benABCDE operon to levels high enough to allow growth on benzoate (12).
A. benzoate → catechol → cis,cis-muconate → succinyl-CoA + acetyl-CoA

B. benPK, benM, benABCDE, catA, ORF 1,2, catM, catBCIJFD

Regulated equally by BenM and CatM
Regulated primarily by BenM
Regulated equally by BenM and CatM
Regulated primarily by CatM
V158M occurred in the effector binding domains (domain I) of both proteins, residues 90 - 303 or 304 in CatM or BenM, respectively. The effector binding domains are amenable to crystallization, whereas the full-length proteins are not. The altered regulatory roles of these variants made it intriguing to understand structural differences between these variants and the wild-type proteins. Since CatM(R156H) and CatM(V158M) can substitute for BenM in the activation of the *ben* genes to enable growth on benzoate, the structures of the effector binding domains of these variants were compared to previously solved wild-type BenM- and CatM-EBD structures. In addition to these analyses, the inducer independent activation of the *ben* and *cat* genes by BenM(R156H/T157S) and CatM(R156H) respectively made it important to compare them to their wild-type counterparts bound to effectors. Although the structure of CatM(R156H)-EBD was solved with muconate, BenM(R156H/T157S) and CatM(V158M)-EBDs were solved in the absence of any effectors.

**MATERIALS AND METHODS**

**Cloning, expression and purification of CatM and BenM-EBD variants**

DNA (*benM5332* allele) encoding the BenM(R156H/T157S) (residues 1 - 304) was PCR amplified from pBAC668 template and cloned into pCR2.1 TOPO, resulting in plasmid pBAC692. The amplification primers were the same ones that were used to amplify the DNA encoding the wild-type BenM-EBD (7) or BenM-full-length (2). The amplification primers contained *NdeI* and *XhoI* restriction sites. This allowed the *benM5332* fragment to be excised from pBAC692 by digestion with *NdeI* and *XhoI* and ligated into a similarly digested pET-21b (novagen) to form pBAC697. Using pBAC697
as a template, DNA encoding just the putative effector binding domain (residues 81-304) was PCR amplified and cloned into a pCR2.1 TOPO vector generating plasmid pBAC758. At sites introduced by the amplification primers, the DNA encoding the BenM(R156H/T157S)-EBD was excised from pBAC758 by digestion with NdeI and XhoI and ligated into similarly digested pET21b (Novagen) to create pBAC698. DNA (catM5153 allele) encoding the CatM(V158M)-EBD (residues 81 - 303) was PCR amplified from pBAC272 template and cloned into pCR2.1 TOPO, resulting in plasmid pBAC678. For the amplification of DNA encoding the CatM-EBD variants, the following set of primers were used: 5'-CATATACATATG AACGGATTTGCAACG-3' (catM-IBD-Nde) and 5'-TCAATTCTCGAGTTCGATGAGTGGCCTGATG-3' (CatM-3-Xho). At sites introduced by the amplification primers, the fragment was excised from pBAC678 by digestion with NdeI and XhoI and ligated into similarly digested pET-21b to create pBAC682. The construction of the expression plasmid for CatM(R156H)-EBD was performed similarly to that of CatM(V158M)-EBD. DNA (catM3102 allele) encoding the putative effector binding domain of CatM(R156H) (residues 81 - 303) was PCR amplified from pIB17 template and cloned into pCR2.1 TOPO to form pBAC680. At sites introduced by the amplification primers, the catM3102 allele fragment was excised from pBAC680 by digestion with NdeI and XhoI and ligated into similarly digested pET-21b to create pBAC683. Cloning into pET-21b translationally fuses a hexa-histidine purification tag to the C-terminus of the protein. Expression plasmids for the full-length versions of CatM(R156H) or CatM(V158M) were constructed similarly to the expression plasmids of the CatM-EBD variants. However, different sets of amplification primers
were employed that amplify the complete sequence. Primers for the amplification of DNA encoding the full-length CatM variants were:

5'-TCAATTCAATATGGAAACTAAGA CACCTCAGA-3' (CatM-5-full-Nde) and 5'-TCAATTCTCGAGTTCTGAGTGGGCTGATATG-3' (CatM-3-Xho). All the plasmids including that of expression plasmids for the full-length versions are listed in table 4.1. The effector binding domains of BenM(R156H/T157S), CatM(V158M), and CatM(R156H) were purified from cells harboring pBAC698, pBAC682, pBAC683, respectively. For expression, the plasmids were transformed into E. coli BL21-Gold (DE3) (stratagene) cells. The expression and purification of these proteins were similar to the procedures employed for the expression and purification of the wild-type BenM- and CatM-EBD proteins as previously described in chapter 3.

**Crystallization of CatM and BenM-EBD variants and effector complexes**

Initial high-throughput crystallization screens (HTS) were done at the Hauptman-Woodward institute using microbatch-under-oil experiments at 298K. The sitting drop method was used to determine initial crystallization conditions (3). This was done by surveying a large (1536 crystallization precipitants) and diverse number of chemical and physical conditions to pinpoint crystallization leads (17). CatM(R156H) - and BenM(R156H/T157S)-EBDs were purified within one week of the HTS and shipped on ice. CatM(V158M)-EBD was purified within two weeks of the HTS. Prior to HTS, BenM(R156H/T157S)-EBD samples were dialyzed as described for BenM-EBDs and concentrated to 14 mg/ml. The CatM(V158M) - and CatM(R156H)-EBD samples were dialyzed as described for CatM-EBD (chapter 3) and concentrated to 11 mg/ml and 12 mg/ml, respectively.
Table 4.1. List of plasmids used in this study\textsuperscript{a}

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Relevant characteristics</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC19</td>
<td>Ap\textsuperscript{r}; PCR cloning vector</td>
<td>(12)</td>
</tr>
<tr>
<td>pCR2.1-TOPO</td>
<td>Ap\textsuperscript{r}; cloning vector</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pET-21b</td>
<td>Ap\textsuperscript{r}; T7 expression vector</td>
<td>Novagen</td>
</tr>
<tr>
<td>pBAC272</td>
<td>Ap\textsuperscript{r}; \textit{cat}M5153 (9819-14225)\textsuperscript{b} in pUC19</td>
<td>(20)</td>
</tr>
<tr>
<td>pIB17</td>
<td>Ap\textsuperscript{r} \textit{cat}M3102 (11950 – 13205)\textsuperscript{b} in pUC19</td>
<td>(12)</td>
</tr>
<tr>
<td>pBAC668</td>
<td>Ap\textsuperscript{r} \textit{ben}M5332 (562 – 3300)\textsuperscript{b} in pUC19</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC677</td>
<td>Ap\textsuperscript{r}; Km\textsuperscript{r}, PCR fragment of \textit{cat}M5153 \textit{allele} (12119-13027)\textsuperscript{b} from pBAC272 in pCR2.1-TOPO</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC678</td>
<td>Ap\textsuperscript{r}; Km\textsuperscript{r}, PCR fragment of \textit{cat}M5153 \textit{allele} (12119-12787)\textsuperscript{b} from pBAC272 in pCR2.1-TOPO</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC679</td>
<td>Ap\textsuperscript{r}; Km\textsuperscript{r}, PCR fragment of \textit{cat}M3102 \textit{allele} (12119-13027)\textsuperscript{b} from pIB17 in pCR2.1-TOPO</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC680</td>
<td>Ap\textsuperscript{r}; Km\textsuperscript{r}, PCR fragment of \textit{cat}M3102 \textit{allele} (12119-12787)\textsuperscript{b} from pIB17 in pCR2.1-TOPO</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC681</td>
<td>Ap\textsuperscript{r}, \textit{cat}M5153 \textit{NdeI}\textsuperscript{c}-XhoI\textsuperscript{c} fragment (12119-13027)\textsuperscript{b} from pBAC677 in pET-21b</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC682</td>
<td>Ap\textsuperscript{r}, \textit{cat}M5153 \textit{NdeI}\textsuperscript{c}-XhoI\textsuperscript{c} fragment (12119-12787)\textsuperscript{b} from pBAC678 in pET-21b</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC683</td>
<td>Ap\textsuperscript{r}, 0.67kb \textit{cat}M3102 \textit{NdeI}\textsuperscript{c}-XhoI\textsuperscript{c} fragment (12119-12787)\textsuperscript{b} from pBAC679 in pET-21b</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC684</td>
<td>Ap\textsuperscript{r}, 0.91kb \textit{cat}M3102 \textit{NdeI}\textsuperscript{c}-XhoI\textsuperscript{c} fragment (12119-13027)\textsuperscript{b} from pBAC679 in pET-21b</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC692</td>
<td>Ap\textsuperscript{r}; Km\textsuperscript{r}, PCR fragment of \textit{ben}M5332 \textit{allele} (1457 -2368)\textsuperscript{b} from pBAC668 in pCR2.1-TOPO</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC697</td>
<td>Ap\textsuperscript{r}, \textit{ben}M5332 \textit{NdeI}\textsuperscript{c}-XhoI\textsuperscript{c} fragment (1457-2368)\textsuperscript{b} from pBAC692 in pET-21b</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC758</td>
<td>Ap\textsuperscript{r}; Km\textsuperscript{r}, PCR fragment of \textit{ben}M5332 \textit{allele} (1457 -2368)\textsuperscript{b} from pBAC697 in pCR2.1-TOPO</td>
<td>This study</td>
</tr>
<tr>
<td>pBAC698</td>
<td>Ap\textsuperscript{r}, \textit{ben}M5332 \textit{NdeI}\textsuperscript{c}-XhoI\textsuperscript{c} fragment (1457-2128)\textsuperscript{b} from pBAC758 in pET-21b</td>
<td>This study</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Ap\textsuperscript{r}, ampicillin resistant; Tcr, tetracycline resistant; Km\textsuperscript{r}, kanamycin resistant.

\textsuperscript{b}Position in the \textit{ben-cat} sequence in GenBank entry (accession number AF009224)

\textsuperscript{c}Restriction sites added in primers used to generate PCR-amplified fragment for cloning
Although the dialysis buffer for BenM-EBD variant contained no imidazole, the dialysis buffers of the CatM-EBD variants required imidazole to keep the proteins in solution. Proteins were concentrated using an ultrafree S-10 centrifuge concentrator (Millipore) and stored at 4 °C until samples were sent for the HTS or set up for crystallization.

The HTS was done by combining 0.2 µl of each protein with 0.2 µl of each precipitating solution per well. Crystallizations were monitored by photography at 1, 7, 14, 21 and 28 days. Although CatM(V158M)-EBD was evaluated without any effectors, CatM(R156H)-EBD was evaluated in the absence and presence of muconate (100 mM), while BenM(R156H/T157S)-EBD was evaluated in the presence and absence of benzoate (100 mM).

The conditions that produced crystals in HTS were repeated in-house. A total volume of 4 µl was used with various ratios of protein to precipitant. Proteins were centrifuged for 5 min at 16000 x g and allowed to cool to room temperature before crystallization trials. Crystals of all the variants were grown using the microbatch under oil method at 15 ºC or 25 ºC. Before cryofreezing in liquid nitrogen for data collection, the crystals were soaked in solutions containing cryoprotectant, usually 35% glycerol plus the final concentration of the conditions from which the crystals grew. Frozen crystals were transported to synchrotron beamlines in dry shippers and transferred to the cryostream operating at 100K.
X-ray analyses and data collection

Diffraction data for the BenM(R156H/T157S)- and CatM(R156H)-EBDs were collected at the Southeast Regional Collaborative Access Team (SER-CAT) at the 22-ID or 22-BM beamlines at the Advanced Photon Source, Argonne, Il with 0.5° oscillations and a wavelength of 1.0 Å. The CatM(V158M)-EBD data set was collected at the Structural Biology Center Collaborative Access Team (SBC-CAT) 19-BM beamline using 0.5° oscillations and a wavelength of 1.00727 Å. Data for CatM(V158M)- and BenM(R156H/T157S)-EBDs were processed and scaled with the beamline versions of HKL2000 (21). The data for both the BenM(R156H/T157S), CatM(V158M) and CatM(R156H) resulted in 99.62%, 99.95%, 94.24% and 99.40% complete data sets respectively (Table 4.2). The space groups for the two BenM(R156H/T157S) variant crystals were found to be both P2_12_12_. The CatM(V158M)- and CatM(R156H)-EBDs belong to spacegroups P_1 and P2_12_2, respectively.

The structures of the two BenM(R156H/T157S)-EBD variants were solved by molecular replacement using coordinates of the previously solved BenM-EBD wild-type structure (Protein Data Bank, accession ID 2F7A) as a molecular replacement model using the program MolRep (24) in the CCP4 suite (1). Since the original molecular replacement model had two copies of the monomer per asymmetric unit, both monomers were separately used as independent search models for both the BenM-EBD variant structures. The CatM-EBD wild-type structure (Protein Data Bank, accession ID 2F7B) was used as the molecular replacement model for the CatM(V158M)- and CatM(R156H)-EBD structures. Rigid body refinement of each initial molecular replacement solution was performed with REFMAC with the subdomains I (residues 88-161; 268-304) and II.
<table>
<thead>
<tr>
<th>Crystallographic Data Collection and Refinement Statistics&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystallization cocktail</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>X-ray source (beamline)&lt;sup&gt;b&lt;/sup&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Space group</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Unit-cell parameters</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Resolution (Å )</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>I/σ(I)</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;merge&lt;/sub&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Resolution range (Å )</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>No. of reflections</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;cryst (%)&lt;/sub&gt;&lt;sup&gt;d&lt;/sup&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;free (%)&lt;/sub&gt;&lt;sup&gt;e&lt;/sup&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>FOM&lt;sup&gt;e&lt;/sup&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Copies in A.U</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>No. of atoms in refinement</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>No. of solvent atoms</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>No. of sulfate ions</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>No. of acetate ions</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>No. of chloride ions</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>No. of glycerol molecules</td>
</tr>
<tr>
<td>Mean temperature factor B (Å²)</td>
</tr>
<tr>
<td>Deviations from ideal geometry</td>
</tr>
<tr>
<td>bond lengths (Å)</td>
</tr>
<tr>
<td>bond angles (°)</td>
</tr>
<tr>
<td>Chirality</td>
</tr>
<tr>
<td>Ramachandran plot statistics (%)</td>
</tr>
<tr>
<td>Most favored</td>
</tr>
<tr>
<td>Additionally allowed</td>
</tr>
<tr>
<td>Generously allowed/disallowed</td>
</tr>
</tbody>
</table>

* Numbers in parentheses refer to the highest resolution shell
* Data sets were collected at the Advanced Photon Source, Argonne, IL. Beamlines 22ID and 22BM are operated by the SouthEast Regional Collaboratory Access Team (SER-CAT) and beamline 19BM is operated by Structural Biology Consortium Collaboratory Access Team (SBC-CAT)
* figure of merit
* $R_{cryst} = \frac{\sum_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)|}{\sum_{hkl} F_{obs}(hkl)}$ for reflections in the working data set.
* $R_{free}$ is the $R$ value for the 5% of the reflections in the cross-validation data set that were not used in refinement.
* The ramanchandran plot was generated with PROCHECK (14)
(residues 162-267) refined as separate bodies. 5% of all measured reflections were randomly selected for use in the calculation of \( R_{\text{free}} \). About fifty rounds of positional and isotropic B-factor refinement using REFMAC (19) and manual modeling using COOT (11, 25) were performed on each of the various models. ARP water identification was also included in the refinement process (23). The refinement process also included TLS temperature factor refinement (26) with the two domains of each monomer refined as separate groups for the CatM(R156H)-EBD crystal. Group B TLS domains for the remaining structures were defined using the TLSMD web server (22).

**Structure analyses**

The quality of the models was analyzed using PROCHECK (15). Molprobity was used to correct the side chains of asparagine, glutamine and histidine residues as well as to identify likely remaining errors during the refinement process (16). The average B-values were calculated using the \( B_{\text{average}} \) program of the CCP4 suite (1). Molecular graphics were drawn with pyMOL (10) or Swiss-PDBviewer (13).

**RESULTS**

**Crystallization of CatM- and BenM-EBD variants**

Video images from high-throughput crystallization screens (HTS) performed by the Hauptman-Woodward institute revealed various conditions that yielded crystals in the absence and presence of effectors. The conditions tended to be completely dependent on the variant proteins and the presence of effectors with few overlaps among the various protein samples. For the CatM(V158M)-EBD, 20 conditions (out of 1536) produced
visible crystals in the absence of effectors. However, the addition of muconate inhibited
crystal formation in the HTS. The crystallization conditions for CatM(V158M)-EBD
were clustered around 40% (w/v) PEG 20000 or PEG 8000 with ammonium or lithium
chloride. In contrast, the HTS screens for CatM(R156H)-EBD showed that the addition
of 100 mM muconate dramatically increased the occurrence of visible crystals. The
favorable response to muconate was not observed for wild-type BenM- and CatM-EBD
crystallization with effectors (5). Effectors inhibited the formation of CatM- and BenM-
EBD crystals as was observed for the CatM(V158M)-EBD variant. The presence of
ammonium sulfate in the crystallization conditions favored the growth of crystals for the
CatM(R156H)-EBD with muconate. The BenM(R156H/T157S)-EBD had the highest
number of conditions that showed crystals (130). Most of these conditions were identical
to those obtained for wild-type BenM-EBD (5). The crystal morphologies of the
BenM(R156H/T157S)-EBD crystals were mostly needle-like (figures 4.2A and B).
Crystals of CatM(R156H)-EBD had more isotropic growth as shown in figure 4.2D. The
conditions obtained from the HTS were successfully reproduced in-house by microbatch
under-oil at room temperature. All crystals grew within one to two weeks with the
exception of CatM(V158M)-EBD, which took about three weeks to reach suitable size
for X-ray diffraction studies. Crystals of CatM(V158M)-EBD grew from the
crystallization condition: 0.1 M lithium chloride, 0.1 M MES, pH 6 and 40% (w/v) PEG
8000. It was observed that larger crystals of CatM(V158M)-EBD were generated when
precipitant and protein were mixed in an eppendorf tube before adding the mixture to the
wells. Crystals of BenM(R156H/T157S)-EBD grew from two different conditions.
Crystal form A grew from 0.015 Magnesium acetate, 0.05 M sodium cacodylate, 1.7 M
**FIG. 4.2** Crystal morphologies of CatM- and BenM-EBD variants.

BenM(R156H/T157S)-EBD crystals (A and B) are identical to the types of crystals previously solved BenM-EBD structures (5). The crystals of CatM(V158M)-EBD (C) and CatM(R156H)-EBD (D) have slightly different morphologies compared to the BenM(R156H/T157S)-EBD crystals. The addition of 100 mM muconate favors the formation of isotropic CatM(R156H)-EBD crystals (D).
ammonium sulfate pH 6. Crystal form B grew from 2.0 M ammonium sulfate. Both crystal morphologies were identical to those of the BenM-EBD crystals and diffracted to resolutions of 1.9Å and 1.8 Å, respectively. Crystals of CatM(R156H) grew from 1.6 M ammonium sulfate, 0.1 M citric acid pH 4 and 100 mM muconate and diffracted to a resolution of 2.5 Å.

Structural determination of CatM(V158M)-, CatM(R156H)- and BenM(R156H/T157S)-EBDs

The variant proteins in their full-length form have shown different phenotypes from their wild-type counterparts in modulating transcriptional regulation (12, 20). It was therefore important to understand the basis of these changes by solving the structures of these variants. Structures of BenM- and CatM-EBDs with and without effectors were reported in chapter 3. The structural studies of full-length versions of the proteins have continually been hampered due to their low solubility in biological buffers. Unfortunately, the full-length variant proteins displayed similar low solubility characteristics that made the wild-type difficult to crystallize. By removing the DNA binding domains and maintaining the purified protein in high concentrations of NaCl (0.5 M) and glycerol, the truncated versions of these variant proteins were adequately soluble and amenable to structure determination studies.

With the exception of the relatively lower resolution CatM(R156H)-EBD structure, all the structures had good refinement statistics ($R_{\text{factors}} < 18\%$ and $R_{\text{free}} < 24\%$). CatM(R156H)-EBD had acceptable $R_{\text{factor}}$ and $R_{\text{free}}$ of 22.6% and 29.4%, respectively. Model and refinement statistics are given in Table 4.2. Phasing of the structures was
performed by using the previous atomic structures of BenM- or CatM-EBDs as molecular replacement models. The refined models had acceptable overall geometry with r.m.s deviations for bond length distances of $\leq 0.008$ Å for all structures, while the r.m.s deviation for bond angle of $\leq 1.16$ for all structures. The Ramachandran phi/psi statistics showed that more than 91% of the dihedral angles were found to be in the most favored region for all the structures.

**Conformational variations of BenM(R156H/T157S) with respect to BenM-EBD**

A mutation was generated in the gene encoding BenM, during the construction of the R156H variant, substituting a serine for threonine at positions 157 (9). BenM(R156H/T157S) activates high level muconate/benzoate-independent benA transcription. The variant is thus in a somewhat semi-activated state in the absence of any effectors. The premise for the engineered mutations in BenM were taken from results of a similar spontaneous mutation in CatM(R156H), which allows high level catB gene expression in the absence of muconate (12, 20). These residues, R156, T157 (in BenM) are located on $\beta$-strand 4 that is part of the proposed tetrameric interface of BenM (Figure 4.3) (chapter 5).

The rationale for elucidating the structures of both the BenM(R156H/T157S) and CatM(R156H)-EBD proteins was to evaluate the mechanism by which these amino-acid substitutions could alter transcriptional activation with respect to the wild-types. The two BenM(R156H/T157S) variant structure forms A and B were nearly identical with 0.14 Å r.m.s deviation. The elucidation of two BenM(R156H/T157S) structures from different crystallization conditions allowed a more accurate comparative analyses with the wild-
FIG. 4.3 Stereo representation of part of the proposed tetrameric interface of BenM taken from the BenM-EBD structure (PDB I.D: 2F8D) (chapter 5). Residues involved in interactions are shown in ball-and-stick representation. The ribbons are shown in transparency. One monomer from a dimer is shown with the backbone and carbon atoms colored purple and the other symmetry related interacting monomer is colored yellow.
type BenM-EBD structures. Previous studies on the interactions of BenM-EBD and CatM-EBD with effectors had shown that muconate binding is enhanced by appropriate helical dipole moments (chapter 3). The binding of muconate to BenM elicits the movement of main chain and side chain atoms in helices by a few angstroms toward the effector. Similar conformational changes as seen in BenM-EBD in response to muconate were observed for the BenM-EBD variants (figure 4.4A and B). The most dramatic structural changes were noticed in α-helices 3, 6 and 7. There is also a minor shift of the β-strand (β4), which contains the residues R156 and T157. All these structural rearrangements were identical in both BenM-EBD variant structures. These changes were also seen in the BenM-EBD structure with bound muconate (Figure 4.4B). The BenM-EBD structure (PDB I.D: 2F7A) has muconate bound to a single monomer (chain B). In this chain the helices of both the variant and the wild-type are in the same location (Figure 4.4B). However, in the other chain that contains no muconate, the difference between the BenM(R156H/T157S) and BenM-EBD is more pronounced especially in α-helix 3. All the conformational changes in the variants or BenM-EBD bound to muconate are towards the centre of the molecule in such a fashion as to create a contracted structure.

**Molecular contacts and surface area changes elicited by the mutations**

In the BenM variant, the change from threonine to serine seems not to have any dramatic local impact on the surrounding residues (Figure 4.5). However, the change from arginine to histidine causes major reorientation of the local residues and hydrogen bonds
FIG. 4.4 Ribbon diagram of BenM(R156H/T157S) superimposed over BenM-EBD (PDB code: 2F6G) (A) or BenM-EBD with bound muconate (B). Approximate locations of 156 and 157 are indicated by arrows (A). The N and C termini are labeled. Also labeled are important secondary structures that undergo changes in conformation. The BenM(R156H/T157S)-EBD structure is colored yellow while the BenM-EBD and BenM-EBD bound to muconate are colored blue in panels (A) and (B) respectively. Muconate is bound to only one of the two monomers in the BenM-EBD-muconate structure. All waters and ions have been removed. The ribbons are rendered in transparency for ease of comparison.
FIG 4.5. Stereo representation of the residues adjacent to the location of the mutation in BenM-EBD (A) and BenM(R156H/T157S)-EBD (B). Hydrogen bonds are shown in broken black lines. An overlay of BenM-EBD (colored grey) and BenM(156H/T157S)-EBD (colored yellow) emphasizes the changes in residue orientations in the beta and loop regions (C).
(Figures 4.5A and B). Although this substitution is one basic amino acid for another basic amino acid, the change alters the molecular interactions associated with the region of the substitution. Residues such as R156 are involved in molecular interactions within the proposed tetrameric interface as presented in chapter 5 and shown in figure 4.3. Analogous residues make up the tetrameric interface for CbnR (18). The alterations in the surface negative and positive charges also cause residues around it to be differently oriented with respect to the wild-type (Figure 4.5). For instance, in all previously solved BenM-EBD structures, R156 is oriented in such a way that it points away from the loop that connects α-helix 3 and β4. However, the variant structures show that the histidine in both monomers is pointed towards the loop. These changes also create an altered hydrogen bonding network between the loop and the β4 in such a way that the loop is more stable and contracts towards the center of the protein.

In BenM- or CatM-EBD structures bound to muconate, T128 is at the end of α-helix 3 and interacts directly with muconate. In the absence of muconate T128 interacts with R156 through water mediated hydrogen bonds (Figure 4.5A). When muconate is bound, there are no interactions between T128 and R156 as the muconate draws α-helix 3 towards itself thus minimizing the contacts between T128 and R156. Similarly, the R156H substitution alters the molecular contacts in such a way that there are no interactions between H156 and T128, instead H156 interacts with T132 of α-helix 3 through water mediated interactions (Figure 4.5B). These substitutions also create a reorientation of residues in the loop regions close to β4 as shown in Figure 4.5C.

A significant difference between the BenM-EBD and the variant structures is the alteration in negative and positive charges at the tetrameric interface due to the R156H
The R156H substitution reduces the negative charge associated with this location thus allowing K148 on the loop to orient towards the beta strand (β4). K148 is also directly involved in molecular interactions within the tetrameric interface. The re-orientation of the side-chain of K148 seems significant as none of the BenM structures solved to date has K148 in this orientation even in the absence sulfates. The substitution causes the side chain of K148 to move towards the surface in the BenM(R156H/T157S) structure form B (Figure 4.5C). This shift causes an increase in the positive charge on the surface thus allowing a sulfate molecule to bind in the form A crystal structure of BenM(R156H/T157S) (Figure 4.6C). The change in surface charges of the tetrameric interface would affect the way the variant protein tetramerizes. This could in turn affect its global structure.

**Comparative studies of the CatM(V158M)- and CatM(R156H)-EBDs**

Although BenM and CatM are 59% identical in sequence, the majority of this identity is contributed by the DNA binding domain which is 85% identical and 98% similar in aminoacid sequence. In comparison, the effector binding domains of both proteins are 53% identical. Although no structure of the DNA binding domain of either CatM or BenM has been obtained, it is plausible to suggest that differences in both protein structures will likely occur in the EBD rather than the DNA binding domain. Overall, the secondary structures of both proteins are similar with the exception of an additional $3_{10}$ helix in BenM-EBD. Despite the structural similarity between both proteins, it is surprising that CatM is not sufficient to replace BenM to allow growth on benzoate but instead variants such as CatM(V158M) and CatM(R156H) are capable.
FIG. 4.6. Charge re-distribution of the molecular surface due to mutations in the BenM(R156H/T157S)-EBD structure. Residues that bring about these charges are labeled. The comparison was done on chain A of each dimeric structure. The BenM-EBD structure shown is shown in panel A. Those of the BenM(R156H/T157S) forms A and B are shown in panels C and B, respectively. The BenM(R156H/T157S) form A structure had a sulfate bound to His156 region. The sulfur atoms of the sulfate residue are shown as yellow colors. The red and blue colors refer to negative and positive charges respectively. The molecular surface was computed using swiss-PDB viewer (13).
The X-ray structure of CatM(V158M)-EBD contained two monomers in the asymmetric unit. When compared to the CatM-EBD structure or the BenM-EBD structure, there are few changes in the structure of the variant (Figure 4.7A). In addition to the large scale similarities between the mutant and wild-type structures, there is very little change found around the mutation or the muconate binding site of the protein. The corresponding region of residue 158 in BenM is a leucine. The V158M substitution is a hydrophobic amino acid for another hydrophobic residue of higher molecular weight. Although this mutation had little effect on the overall structure, it definitely disrupts the positive charge distribution locally. As shown in figure 4.8B, the valine residue at 158 of CatM-EBD accommodates the binding of a sulfate. Sulfates would bind to regions of the protein with a net positive charge. However, when the valine residue is changed to a much larger hydrophobic residue like methionine in CatM, the positive charge at this region is disrupted. The corresponding residue in another LysR-type transcriptional regulator, CbnR, is located near the tetrameric interface (18). It is possible that changes in interface charge around this region might subtly alter the tetrameric interactions within the interface, thus producing similar surface charges to that of BenM which has a leucine in the similar position (Figure 4.8B). These changes might alter the way this variant tetramerizes, thus affecting its regulatory role. This structure of CatM(R156H)-EBD contained four monomers in the asymmetric unit. The crystals grew out of precipitation solution containing 0.1 M muconate. Electron densities representing muconate were identified in all four monomers. The location of the bound muconate was identical to that seen in previously solved BenM and CatM structures bound to muconate with similar hydrogen bond interactions. The four monomers were nearly identical with r.m.s
FIG. 4.7. Comparison of structures of BenM, CatM and their variant EBDs. An overlay of CatM(V158M)-EBD and CatM-EBD colored blue and yellow, respectively (A). Overlay of BenM-EBD bound to muconate and CatM(R156H)-EBD bound to muconate colored blue and yellow, respectively (B). Overlay of CatM(V158M)-EBD and CatM(R156H)-EBD bound to muconate colored blue and yellow respectively (c).
FIG. 4.8. Stereo representation of an overlay of CatM-EBD, CatM-EBD bound to muconate and CatM(R156H)-EBD bound to muconate colored pink, green and yellow respectively (A). The residues and secondary structures surrounding the R156/V158 residues are shown in ball-and-stick representation. The secondary structures are shown in transparency for ease of comparison. Stereo representations of an overlay of CatM-EBD, BenM-EBD and CatM(V158M)-EBD (B), backbone atoms are colored yellow, green and gray respectively. The sulfate molecule is from the CatM-EBD structure. Figures were drawn with pymol (10).
deviations of about 1.0 Å with the exception of chain B which had a r.m.s deviation of greater than 2.5 Å when compared to other three monomers in the asymmetric unit. Interestingly, this chain had an average B value of 64.4 compared to the other chains A, C and D which had average B-values of 75.3, 74.8 and 83.4 respectively. All the monomers had a contracted structure similar to the BenM(R156H/T157S)-EBD structures. As shown in figure 4.7B and C, αH3 and 7 shift towards the center of the dimer. There are also shifts in the β4 and the loop between αH3 (residues 128 - 137) and β4 in such a way that the loop (residues 125 and 127) is drawn closer to the muconate moiety compared to CatM(V158M)- and CatM-EBD. The most dramatic change between the CatM(R156H) and all the other structures was seen in α-helix 3 (figure 4.7B and C).

**DISCUSSION**

**Mutations modify the tetrameric interface**

The primary goal of this study was to establish the structural basis for the altered regulatory roles of the variant proteins BenM(R156H/T157), CatM(R156H) and CatM(V158M) in controlling benzoate degradation. Amino acid substitutions that affect inducer response or lead to constitutive activity of LTTRs are often located in the tetrameric interface and/or effector binding pocket (4, 18). The structure of CbnR showed that residues mediating effector response were mostly located near the tetrameric interface. In another LTTR, OxyR, it was shown that the tetrameric orientations and interactions were different between the oxidized and reduced states of the protein (4). The fact that these spontaneous mutations in CatM and BenM are clustered in this region is therefore not surprising. These mutations enabled CatM to serve the sole regulator of the
ben and cat genes. In addition, CatM(R156H) enabled CatM-mediated activation of the

cat genes in the absence of inducer. The change in residue 158 from valine to methionine

may seem conserved as it is a substitution from one hydrophobic residue to another

hydrophobic residue. However, the structures showed that the change disrupts the

negative charge associated with the surface (figure 4.8B). The equivalent position in

BenM is a hydrophobic amino acid, leucine, which has a similar molecular weight to

methionine. The structural similarity of CatM(V158M)-EBD and CatM-EBD was not

surprising as the physiological assays conducted for gene regulation by both proteins

were not dramatically different. CatM(V158M) caused a two fold increase and decrease

in expression of benA and catB, respectively in response to muconate compared to the

CatM wild-type (12). The changes in the tetrameric interface for CatM(V158M)-EBD

protein were not as dramatic as those observed for the CatM(R156H)- or

BenM(R156H/T157S)-EBDs. As seen in the BenM(R156H/T157S)-EBD structures,

amino acid substitutions in this site alter the local orientation of neighboring residues

especially K148, thus modifying the surface charges which might affect tetrameric

interactions. A change in tetrameric interactions could affect the manner by which this

protein tetramerizes, thus affecting their regulatory roles. Interestingly, residues R156

and K148 play major roles in the tetramerization and oligomerization of BenM-EBD

structures (chapter 5).

Structural basis for altered functional properties

The CatM(R156H) and BenM(R156H/T157S) showed changes in conformation from

the wild-type structures (Figures 4.4 and 4.7). The low r.m.s deviation (0.21 Å for chains
A and B combined) between both BenM(R156H/T157S) independent structures indicate that these changes are significant and are not due to any crystallization artifacts. Secondly, the fact that these changes seem to mimic a BenM-EBD structure bound to muconate supports the physiological studies in which these proteins are in a semi-activated state. Although acetate or sulfates were modeled in the muconate binding sites of the two BenM(R156H/T157S) structures, it is unreasonable to speculate that these compounds could be causing these conformational changes. In a third structure of BenM(R156H/T157S)-EBD that had no molecules bound to the protein (not presented due to poor quality of the data), conformational changes similar to the other two BenM(R156H/T157S)-EBD structures were observed. The average r.m.s deviation between any of the monomers of BenM(R156H/T157S)-EBDs and chain A of BenM-EBD is about 0.9 Å (Table 4.3) However, the r.m.s deviation decreases to 0.7 Å when any of these monomers are compared to chain B of BenM-EBD that has muconate bound to it. This change in reduction of the r.m.s deviation is consistent with these variants mimicking a BenM-EBD structure with bound muconate. The trend is different when monomers of BenM without muconate are compared with those that do have muconate bound to the protein (Table 4.3). In comparing CatM variant structures with the BenM-EBD structures, there is relatively low r.m.s deviations between the monomers of CatMR156H and BenM with bound muconate (Table 4.4). It is possible that this same trend would have been observed for CatM(V158M)-EBD if muconate were bound to the protein.

The analyses of the B-values of the β4 strand in all the BenM/CatM wild-type and variant structures show a general trend (Table 4.5). β4 is the only secondary element that
Table 4.3. Superimposition of monomers of BenM-EBD, BenM-muconate-EBD and BenM(R156H/T157S)-EBD structures

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BenM chain A (2F7A)</td>
<td>0.79</td>
<td>0.49</td>
<td>0.91</td>
<td>0.94</td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>BenM-muconate chain B (2F7A)</td>
<td>0.92</td>
<td>0.99</td>
<td>0.70</td>
<td>0.78</td>
<td>0.74</td>
<td>0.77</td>
</tr>
</tbody>
</table>

The table shows the average r.m.s deviation in atomic positions (Å) after a least squares imposition of the CA atoms of residues 96 - 295 in each monomer (residues 175-180, 277 and 278 were removed from the calculations due to the unusually high r.m.s deviation in these residues between monomers of the same structure). The PDB ID codes for each structure are indicated in bracket.
Table 4.4 Superimposition of various monomers of CatM-EBD, BenM-EBD and CatM-EBD variant structures.

<table>
<thead>
<tr>
<th></th>
<th>CatM-CCM chain A (2F7C)</th>
<th>CatM V158M chain A (2H98)</th>
<th>CatM V158M chain B (2H98)</th>
<th>CatM R156H muconate chain C (2H9Q)</th>
<th>CatM R156H muconate chain D (2H9Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenM chain A (2F6G)</td>
<td>1.53</td>
<td>1.30</td>
<td>1.23</td>
<td>1.28</td>
<td>1.30</td>
</tr>
<tr>
<td>BenM chain B (2F6G)</td>
<td>1.70</td>
<td>1.52</td>
<td>1.46</td>
<td>1.30</td>
<td>1.40</td>
</tr>
<tr>
<td>BenM chain A (2F7A)</td>
<td>1.87</td>
<td>1.70</td>
<td>1.56</td>
<td>1.43</td>
<td>1.51</td>
</tr>
<tr>
<td>BenM-muconate chain B (2F7A)</td>
<td>1.43</td>
<td>1.47</td>
<td>1.66</td>
<td>1.07</td>
<td>1.11</td>
</tr>
</tbody>
</table>

The table shows the r.m.s. deviation in atomic positions (Å) after a least squares imposition of the CA atoms of residues 96 - 295 in each monomer. The PDB ID codes for each structure are indicated in bracket.
Table 4.5. Average B-values of residues on the β4 of wild-type and variant structures of CatM- and BenM-EBD

<table>
<thead>
<tr>
<th>Chain</th>
<th>BenM-EBD</th>
<th>BenM(R156H/T157S)-EBD</th>
<th>BenM(R156H/T157S)-EBD</th>
<th>CatM-EBD (chain A)</th>
<th>CatM(R156H)-EBD (chain B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A153</td>
<td>37.5</td>
<td>26.4</td>
<td>26.2</td>
<td>18.0</td>
<td>73.4</td>
</tr>
<tr>
<td>I154</td>
<td>39.1</td>
<td>26.3</td>
<td>26.3</td>
<td>19.3</td>
<td>73.2</td>
</tr>
<tr>
<td>K155</td>
<td>41.6</td>
<td>26.7</td>
<td>25.5</td>
<td>22.1</td>
<td>76.0</td>
</tr>
<tr>
<td>R156</td>
<td>40.6</td>
<td>22.5</td>
<td>22.5</td>
<td>17.9</td>
<td>75.5</td>
</tr>
<tr>
<td>T157</td>
<td>32.5</td>
<td>20.9</td>
<td>20.6</td>
<td>16.7</td>
<td>69.6</td>
</tr>
<tr>
<td>L158</td>
<td>29.8</td>
<td>19.3</td>
<td>18.8</td>
<td>16.2</td>
<td>66.1</td>
</tr>
<tr>
<td>L159</td>
<td>26.1</td>
<td>16.7</td>
<td>18.0</td>
<td>14.5</td>
<td>63.2</td>
</tr>
<tr>
<td>R160</td>
<td>27.4</td>
<td>20.5</td>
<td>20.7</td>
<td>14.4</td>
<td>65.9</td>
</tr>
<tr>
<td>N161</td>
<td>24.3</td>
<td>20.9</td>
<td>18.2</td>
<td>17.5</td>
<td>68.3</td>
</tr>
<tr>
<td>E162</td>
<td>24.5</td>
<td>19.4</td>
<td>16.3</td>
<td>16.8</td>
<td>71.7</td>
</tr>
<tr>
<td>R163</td>
<td>24.8</td>
<td>24.4</td>
<td>21.1</td>
<td>19.4</td>
<td>69.7</td>
</tr>
<tr>
<td>L164</td>
<td>21.4</td>
<td>16.2</td>
<td>13.4</td>
<td>12.8</td>
<td>65.4</td>
</tr>
<tr>
<td>M165</td>
<td>23.0</td>
<td>17.6</td>
<td>13.8</td>
<td>16.8</td>
<td>65.8</td>
</tr>
<tr>
<td>V166</td>
<td>20.7</td>
<td>14.4</td>
<td>11.8</td>
<td>13.7</td>
<td>62.7</td>
</tr>
<tr>
<td>A167</td>
<td>20.0</td>
<td>12.6</td>
<td>10.4</td>
<td>13.8</td>
<td>61.6</td>
</tr>
<tr>
<td>V168</td>
<td>20.0</td>
<td>12.5</td>
<td>10.4</td>
<td>14.0</td>
<td>60.0</td>
</tr>
<tr>
<td>H169</td>
<td>21.4</td>
<td>13.1</td>
<td>11.4</td>
<td>13.7</td>
<td>60.2</td>
</tr>
<tr>
<td>A170</td>
<td>21.0</td>
<td>12.7</td>
<td>10.9</td>
<td>18.0</td>
<td>57.7</td>
</tr>
</tbody>
</table>

*The average B-values were calculated using the Baverage program of the CCP4 suite.

bThis refers to the average B value for the whole chain. For the BenM and BenM-EBD variant structures, chain A was used. For the CatM(R156H)-EBD, chains A and B were used as indicated. The residues are numbered according to the wild-type BenM-EBD.
fully spans both domains of the effector binding domain. The flexibility of this secondary structure element is far more pronounced in the region of the R156 and then tapers down along the strand as evidenced by the lower B-values. The change was most dramatic in β4 for the CatM(R156H) that had an exceptionally high average B-values. For instance, in chain A, the B-value goes from 73.6 (A153) to 75.5 (H156) and then reduces to 57.7 (K170) at the end of the strand. The flexibility of this region at the beginning of β4 suggests that mutations in this region would have a greater impact on altering molecular interactions as observed for BenM(R156H/T157S) (Figures 4.5A and B). These alterations are then easily amplified and propagated to other parts of the protein.

The substitutions reported here provide examples of how residue changes could alter the functionality of a regulatory protein. The substitution, R156H in CatM and BenM are at significant locations in the protein based on the following reasons: (i) it interacts with residues at the effector binding pocket; (ii) it interacts with residues of α-helix 3 that directly interacts with linker helix that connects DNA binding domain with the effector binding domain (figure 4.9); (iii) it is within the proposed tetrameric interface of the protein; (iv) it is also located in a region of the protein that is very flexible as evidenced by the high temperature factors. Additionally, R156 in some of the BenM- and CatM-EBD structures adopts alternate conformations. The most common conformation of arginine is at a position where it allows for optimized water-mediated hydrogen bond networks to α-helix 3. The other conformation disrupts this network. These two alternate conformations may allow α-helix 3 to move freely in response to muconate. This is important as both proteins could act as repressors or activators and both states are
**FIG 4.9.** Structural model of a compact form of a full-length CatM monomer using CatM-EBD (PDB ID: 2F7B) and CbnR (18) as templates. Residues in the interface of α-helix 3 and the linker helix are shown as sticks. The directional changes in orientation of α-helix 3 as observed for the CatM(R156H)-EBD in relation to the CatM-EBD structure are represented by arrows. Hydrogen bonds are showed as black broken lines.
physiologically essential for balanced regulation of the genes needed for benzoate degradation.

**Structural basis for increased benA expression by CatM(R156H) and CatM(V158M) without BenM**

CatM(R156H) and CatM(V158M) increased benA expression by five- and two-fold respectively, in response to muconate compared to the wild-type CatM (12). Because CatM(R156H)-EBD was solved with muconate and not CatM(V158M)-EBD, we could compare the structure of CatM(R156H)-EBD to those of BenM- and CatM-EBD with bound muconate. In comparing the structures of CatMR(156H)-EBD and CatM-EBD with bound muconate, α-helix 3 was the secondary structure with the highest r.m.s deviation compared to other secondary structural elements. However, the r.m.s deviation in α-helix 3 decreases substantially when CatM(R156H)-EBD is compared to the BenM-EBD monomer with bound muconate. It is possible that α-helix 3 could be an important transmitting signal of how these CatM variant proteins interact with the benA promoter or RNA polymerase. It might not be plausible to draw any concrete conclusions since these structures represent the effector binding domain of these proteins. Nonetheless, a modeled full-length structure of CatM based on a the structure of CatM-EBD and a monomer of CbnR (compact form) (18) shows that the α-helix 3 could form various potential salt bridges and hydrogen bonds with the linker helix that connects the effector binding domain and DNA binding domain (figure 4.9). Shifts in α-helix 3 may alter its interactions with the linker helix or cause actual shifts in the linker helix. The resulting effect could be a re-orientation of the DNA binding domains in such a way as to optimize
the interaction of the CatM variant proteins with RNA polymerase and/or the *benA* promoter in such a way as to increase *benA* expression in the absence of BenM.
REFERENCES


CHAPTER 5

OLIGOMERIZATION OF BENM, A LYSR-TYPE TRANSCRIPTIONAL REGULATOR: STRUCTURAL BASIS FOR THE AGGREGATION OF PROTEINS IN THIS FAMILY¹

¹Obidimma C. Ezezika, Sandra Haddad, Ellen L. Neidle and Cory Momany. To be submitted to Acta Crystallographica Section D
ABSTRACT

LysR-type transcriptional regulators comprise the largest family of homologous regulatory DNA-binding proteins in proteobacteria. A general challenge in the crystallization of this family of proteins has been insolubility and precipitation difficulties when working with high concentrations of full-length versions of these proteins. A general oligomerization scheme is proposed for this family of proteins based on two structures of the effector binding domain of BenM in two different space-groups, P4₃22 and C222₁. These structures used the same oligomerization scheme of dimer-dimer interactions as another LysR-type transcriptional regulator, CbnR, whose full-length structure is available [Muraoka et al., (2003). J. Mol. Biol. 328,555-566]. Evaluation of packing relationships and surface features showed that BenM can form infinite oligomeric arrays in crystals through these dimer-dimer interactions. By extrapolation to the liquid phase, this biologically relevant property likely contributes to the significant difficulties encountered when trying to crystallize members of this family. The oligomerization of dimers to form the biologically important tetramers leaves unsatisfied oligomerization sites that under the right conditions will favor higher-order oligomerization leading to solubility problems. A detailed model by which BenM and other LysR-type transcriptional regulators form these arrays is hereby proposed.
INTRODUCTION

BenM is a transcriptional regulator found in the soil bacterium *Acinetobacter baylyi* ADP1. This regulator together with another transcriptional regulator, CatM, controls a large set of genes needed for aromatic compound degradation in this bacterium (2, 7, 8, 10, 14). BenM is a member of the family of LysR-type transcriptional regulators (LTTR). This family constitutes the largest family of transcriptional regulators in proteobacteria (16, 29, 31). The roles of genes controlled by this LTTR family are diverse and include: synthesis of virulence factors, CO$_2$ fixation, antibiotic resistance, catabolism of aromatic compounds, nodule formation of N$_2$ fixing bacteria, and amino acid biosynthesis. LTTRs are recognized in abundant and diverse bacterial genera (12). BenM specifically belongs to a subclass of this family involved in aromatic compound catabolism (34).

Mutational analyses of this group of regulators demonstrate that the N-terminal region is required for DNA binding. This region comprises residues 1-66 and displays high sequence identity among members (32). Removal of the N-terminal region aids structural studies by circumventing insolubility problems associated with the full-length versions of these proteins as observed in BenM and CatM (7) and other LysR-type regulators such as CysB, OxyR and DntR (6, 33, 35, 37). As a result, the structure determinations of truncated versions of LTTRs have been more successful than for the full-length structures. CbnR is the only example of a full-length LTTR structure (25), although DntR crystallizes with poor electron density for the DNA-binding domain (Smirnova et al., 2004). CbnR monomers have two different conformations of the same polypeptide chain, making the tetrameric molecule a dimer of dimers. BenM exists as a
tetramer in its active, full-length form (3). Tetramers are generally the active form of this family of regulators as noted in CbnR, CysB, NahR and DntR (17, 25, 32, 33).

BenM responds to two effector ligands, benzoate and cis,cis-muconate, synergistically to activate transcription of the ben genes (3). In contrast, CatM, a homolog of BenM, responds only to cis,cis-muconate. Recently, the structures of the effector binding domains of BenM and CatM were determined with and without their effectors (chapter 3). These structures identify two distinct binding sites for the effectors and identify conformational changes associated with ligand binding that are likely associated with transcriptional activation. One unexplored area of the BenM structural studies is the interaction among its subunits to form tetramers and high-order oligomers. Here we report two structures of the BenM-EBD in different space groups that form tetramers and high-order oligomers in their unit cells. The structures revealed a general scheme by which this family of proteins might tetramerize and form high-order oligomers. Surface oligomerization domains were identified and analyzed in the context of other members of this family.

**MATERIALS AND METHODS**

**Purification**

The effector binding domain of BenM (BenM-EBD) was expressed from a pET-21b–based vector (Novagen) in which the BenM-EBD gene was translationally fused to a hexahistidine encoding purification tag at the C-terminus of the protein as previously described (7, 9). Purification of BenM-EBD proteins was performed as previously
described in chapter 3 using a 5 ml Hi-TRAP metal-chelating column (GE Biosciences) charged with Ni$^{2+}$.

**Crystallization and X-ray data collection**

Before crystallization, BenM-EBD was dialysed into buffer consisting of 20 mM Tris-HCl pH 7.9, 0.5 M NaCl and 10% [v/v] glycerol. Crystallization setups were done using the microbatch under oil method (5) at 15 °C using conditions that had been identified in high-throughput screens at the Hauptman-Woodward Institute (24). Two conditions generated crystals having different properties from those identified in earlier reports (7). The two crystals are herein referred to as crystal/structures A and B (Table 5.1). For structure A, 2 µl of protein solution (6 mg ml$^{-1}$) was mixed with 2 µl of crystallization precipitant solution containing 0.1 M LiCl, 0.1 M N-Tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid (TAPS) pH 9, 5 mM cis,cis-muconate (from a 0.2 M stock adjusted to pH 7 with NaOH) and 20 % PEG 8000. Crystal B growth conditions were performed identically except with a different precipitant, in this case 100 mM KBr, 100 mM N-Cyclohexyl-3-aminopropanesulfonic acid (CAPS) pH 10, and 20 % PEG 4000. Proteins were centrifuged briefly before addition of the precipitant in an Eppendorf microcentrifuge ~5 min at 14000 × g. Crystals were transferred into appropriate cryosolvents that contained all of the crystallization components with final amounts of 35 % glycerol or 35 % polyethylene glycol 400 for crystals A and B respectively. Crystals were then flash frozen in liquid N$_2$ and shipped overnight in a dry cryogenic shipper before data collection.
X-ray data collection and structure refinement

Diffraction data were collected at the Structural Biology Center Collaborative Access Team (SBC-CAT) 19-BM beamline at the Advanced Photon Source, Argonne, Il using 0.5° oscillations, a wavelength of 1.00727 Å, and the detector set at 194 mm. Data were processed and scaled with the beamline version of HKL2000 (28) and resulted in 99.9% and 98.4% complete data sets extending to 2.2 Å and 2.7 Å for structures A and B respectively (Table 5.1). The space groups were ultimately found to be P4322 and C2221 for structure structures A and B respectively. Structure A had a monomer in the asymmetric unit. Two monomers in the asymmetric unit were observed for structure B. There were eight monomers in the unit cell of both structures. Structure A was initially indexed and the structure solved in a lower symmetry space group (Laue 4/m) until it was realized that the refined structures of the two monomers were identical except in two poorly defined regions at the N- and C-termini that were modelled differently. The correct space group was P4322. The data were reindexed and scaled in the correct space group, P4322. The C2221 data set, although having a triclinic cell very close to the 4/mmm space group, does not satisfy 4/mmm symmetry. The structure solutions of both A and B were straight forward using coordinates of the previously solved BenM-EBD native structure (Protein Data Bank, accession ID 2F7A) as a molecular replacement model. The program MolRep in the CCP4 suite (1) was used for the molecular replacement. Since the original BenM-EBD structures solved in P212121 had two copies of the monomer per asymmetric unit, both monomers were separately used as independent search models in the BenM-EBD structures. Rigid body refinement in REFMAC with the sub domains I (residues 88-161; 268-304) and II (162-267) refined as
**Table 5.1** Crystal properties, data collection, and refinement statistics. Values in parentheses are for the highest resolution shell.

<table>
<thead>
<tr>
<th></th>
<th>Structure A (2F97)</th>
<th>Structure B (2F8D)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystallization conditions</strong></td>
<td>0.1 M LiCl, 0.1 M TAPS, pH 9, 5 mM cis,cis-muconate, 20 % PEG 8000</td>
<td>100 mM KBr, 0.1 M CAPS, pH 10, 20 % PEG 4000</td>
</tr>
<tr>
<td><strong>Data collection statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P4$_3$22</td>
<td>C222$_1$</td>
</tr>
<tr>
<td>Unit-cell parameters (Å)</td>
<td>$a = 70.0, b = 70.0, c = 187.7$</td>
<td>$a = 94.1, b = 106.4, c = 184.3$</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>70 - 2.2</td>
<td>46.1 - 2.7</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.99 (99.9)</td>
<td>98.4 (98.1)</td>
</tr>
<tr>
<td>$I/\sigma(I)$</td>
<td>65.96 (5.91)</td>
<td>28.1 (5.85)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.09</td>
<td>7.16</td>
</tr>
<tr>
<td>$R_{merge}$ (%)</td>
<td>6.3 (76.9)</td>
<td>9.3 (42.1)</td>
</tr>
<tr>
<td><strong>Refinement statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>46.9-2.2 (2.28-2.20)</td>
<td>46.1 – 2.7 (2.80-2.70)</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>23294 (1626)</td>
<td>24161 (1685)</td>
</tr>
<tr>
<td>$R_{cryst}$ (%)</td>
<td>17.6 (24.2)</td>
<td>14.8 (22.2)</td>
</tr>
<tr>
<td>$R_{free}$ (%)</td>
<td>20.9 (30.6)</td>
<td>19.5 (24.1)</td>
</tr>
<tr>
<td>$N_e$. of solvent atoms</td>
<td>295</td>
<td>596</td>
</tr>
<tr>
<td>Mean temperature factor B (Å$^2$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main chain</td>
<td>38.0</td>
<td>27.7</td>
</tr>
<tr>
<td>Side chain</td>
<td>39.3</td>
<td>27.8</td>
</tr>
<tr>
<td>Solvent</td>
<td>58.4</td>
<td>52.2</td>
</tr>
<tr>
<td>Root-mean-square deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>Bond angles ($)</td>
<td>1.115</td>
<td>1.301</td>
</tr>
<tr>
<td>Ramachandran plot statistics (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored</td>
<td>94.7 (177 residues)</td>
<td>93.4 (356 residues)</td>
</tr>
<tr>
<td>Additionally allowed</td>
<td>5.3 (10 residues)</td>
<td>6.6 (25 residues)</td>
</tr>
<tr>
<td>Generously &amp; disallowed</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
separate bodies were performed before any refinement and model building. 5% of all measured reflections were randomly selected for use in the calculation of Rfree. Several rounds of positional and isotropic B-factor refinement using REFMAC (27, 36) and manual modelling using O (20), Swiss-PDBviewer (15), and COOT (13) with ARP water identification (30) were performed. Molprobity (23) was used to correct the side chains of histidine, asparagine and glutamine residues as well as to identify likely remaining errors during the refinement process. The refinement process included TLS temperature factor refinement (38) with the two domains of each monomer refined as separate groups. 288 water molecules were incorporated in structure A, while 574 water molecules were incorporated in structure B.

Structure analysis

The quality of the models was analyzed using PROCHECK (22). Hydrogen bonds between proteins atoms were calculated using Swiss-PDBviewer (15) with the default parameters for distance and angles. Salt bridges were inferred when aspartic or glutamic acid side-chain carboxyl oxygen atoms were found to be within 4.0 Å of the side chain nitrogen atoms of arginine, lysine, or histidine amino acids. Accessible surface areas and interface characteristics were calculated with a protein-protein interaction server (http://www.biochem.ucl.ac.uk/bsm/PP/server/) based on work done by Jones and Thornton (18, 19). Molecular graphics were drawn with pyMOL (11).
RESULTS AND DISCUSSION

Structure determination

Previous structures of BenM-EBD with and without benzoate and/or cis,cis-muconate crystallized in space group P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1} from conditions having low pH precipitants (pH 4.6) and complex anion compositions (acetate, chloride and sulfate) (7). Crystals grown under the higher pH conditions reported herein, where precipitant solutions were at pH 9 and 10, were found to diffract in two different space groups with different unit cell parameters. While still containing high chloride levels due to the presence of 0.25 M NaCl from the protein buffer, no sulfate or acetate ions were present in the crystallization cocktail. Because the protein solution itself contained significant buffering capacity due to the high imidazole and minimal tris concentrations, the actual pH of the final crystallization conditions would lie between that of the protein solution and the pH of the precipitant solution. Although a general morphology of the dimeric structure and its interaction with inducers was presented in chapter 3, inference on the tetrameric arrangement and possible formation of oligomers was not structurally presented. The structures of the two higher pH crystal forms here provide a novel view of higher-order oligomerization.

Crystal characteristics and data collection statistics for both crystal forms in the P4\textsubscript{3}22 and C222\textsubscript{1} space groups (herein referred to as structures A and B respectively) are given in Table 5.1. Structure A was refined to 2.2 Å resolution with a crystallographic R-factor of 17.6% and Rfree of 20.9%, while structure B was refined to 2.7 Å resolution with a crystallographic R-factor of 14.8% and Rfree of 19.5% (R\textsubscript{free} was calculated on a 5% subset of all measured reflections). Phasing of the structures was straightforward
using the previous atomic structures as molecular replacement models. Model and refinement statistics are given in Table 5.1. The refined models had acceptable overall geometry with r.m.s deviations for bond length distances of 0.008 Å for both structures, while the r.m.s deviations for bond angles were 1.15° and 1.30° for structures A and B, respectively. The Ramachandran phi/psi statistics showed that more than 93% of the dihedral angles were found to be in the most favored region for both structures with none in the generously allowed or disallowed regions. Crystal structures A and B contained a monomer or dimer in the asymmetric unit, respectively. The comparison of both structural models was calculated using the Swiss-PdbViewer (15). The r.m.s deviation between the subunit A of structure A and either subunits A or B of structure B was less than 1.21 Å for the backbone and 1.45 Å for all atoms. When comparing the main-chain trace of the refined structures with the previously solved structures, no significant differences were observed. The overall similarities between these structures and other determined BenM-EBD structures were comparable with r.m.s deviation of less than 1.6 Å for all atoms.

**The monomer-monomer interface of BenM-EBD**

The full-length native BenM protein is a homotetramer in solution (3). This is consistent with results that show that most LysR-type regulators exist as tetramers (25, 34). However, gel-filtration studies carried out on the truncated form of the protein in which the 88-N terminal residues were removed showed that BenM-EBD existed as a homodimer in solution (7). In all of the structures of the BenM-EBD determined, obtained from lower pH in previous work and also here, two monomers are intimately associated with each other to form a stable dimer with a common interface. In several
cases, two monomers are in the asymmetric unit (the low pH space group P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1} and structure B here). The monomeric structure of BenM-EBD is an $\alpha/\beta$ structure that consists of 9 $\alpha$-helices and 9 $\beta$-strands (figure 5.1A). Two subunits form a dimer by anti-parallel, side by side alignment of the monomers, which are related by a non-crystallographic two fold symmetry as observed for structure B (figure 5.1B). Although the number of molecules in the asymmetric unit of structure A was one, the other monomer could be generated by applying the crystallographic symmetry operator $-x, y, -z$ to form a dimer similar to the molecule in the asymmetric unit of structure B. Stable interactions are present between both monomers to form the dimeric arrangement in structure B. These interactions comprise hydrogen bonds, hydrophobic interactions and salt bridges. The interaction between both subunits in the dimer buries a surface of 1315 Å\textsuperscript{2} on each molecule. The core region of the interface is composed of hydrophobic residues: Leu101, Pro106, Ile09, Leu123, Val227, Leu229, Leu233, Ala235, Ala236, Ile250 and Leu252. Surrounding these hydrophobic residues are hydrophilic residues that provide hydrogen bonds and salt bridges: these residues included Arg113, Glu125, Arg225, Glu226, Gln228, Ala236, Glu238, Ser249, and Gln251. These interactions taken together are very similar to the two monomer-monomer interfaces observed for CbnR (26).

The tetramerization interface

Two neighboring dimers in the crystal lattice interact with each other to form a well-defined tetrameric unit. In structure B, a dimer interacted with adjacent subunits in the crystal lattice to form a tetrameric unit (Figure 5.2B). The adjacent dimer was defined
**FIG. 5.1.** Ribbon representation of the monomeric and dimeric structures of BenM-EBD structures A and B. Panel A, shows the structure of a monomer of BenM-EBD. The secondary structures important for tetramerization are indicated. The monomer can be divided into two major domains: domain I (residues 81-161 and 268-296) and domain II (residues 162-267). Panel B, view of the dimeric structure of BenM-EBD. Secondary structures are colored from blue to red going from the N-terminus to C-terminus.
FIG. 5.2. Ribbon representation of BenM-EBD tetramer structure A in the P4_{3}22 space group (Panel A) and structure B tetramer in space group C222_{1} (panel B). Secondary structures of each subunit are colored from blue to red going from the N-terminus to C-terminus.
by the crystallographic symmetry operator: \(-x+1, y, 1/2-z\). For the crystal structure A, which contained a monomer in the asymmetric unit, the crystal lattice showed the interaction of subunits to form a similar dimer that could generate a subsequent tetramer using the symmetry operation \(y, x, 1/4-z\) applied to both monomers. The higher pH of the crystallization conditions could likely explain why BenM-EBD packs as tetramers when compared to the previously solved structures where only dimeric complexes were observed. Such phenomenon is not uncommon (21). The general characteristics governing the tetrameric interface between both subunits were comparable. The surface area buried by the interaction of the dimers was approximately 998 Å² and 883 Å² for structures A and B respectively. This was similar to the 855 Å² observed in CbnR's tetrameric interface (Table 5.2). The planarity was also similar between the BenM-EBD structures and CbnR, which is a measure of how far the interface residues deviate from a best fit plane calculated through the 3-dimensional co-ordinates of the atoms in the interface using principal component analysis (19). Other parameters such as the length and breadth of the interfaces, length/breadth ratios and relative percentages of polar and non-polar atoms in the interface were similar between the BenM structures and CbnR. “Alpha” was the predominant secondary classification of the interface with the involvement of some random coil and beta strand residues. The hydrophobic residues in the interface consisted of Leu147, Lys148, Ile149, Leu184, Pro201, Gly215, Leu216, Pro263 and Pro268. Other residues, some of which were involved in hydrogen bonding and/or the formation of salt bridges included Thr128, Lys129, Lys148, Ser150, Asp151 Arg156, His183, Asn185, Asp186, Thr205, His206, Asn209, Ser212 Asp213, His214, Gly215, Asp262 and Asp264.
Table 5.2 Analysis of the tetrameric interfaces of the two BenM-EBD structures with the structure of CbnR from *Ralstonia euphora*

<table>
<thead>
<tr>
<th></th>
<th>Structure A (2F97)</th>
<th>Structure B (2F8D)</th>
<th>CbnR (1IZ1)(^{a})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface accessible surface area, Å(^2)</td>
<td>882.84</td>
<td>997.6</td>
<td>854.57</td>
</tr>
<tr>
<td>% Interface accessible surface area</td>
<td>8.01</td>
<td>8.79</td>
<td>8.17</td>
</tr>
<tr>
<td>Planarity(^{b})</td>
<td>2.78</td>
<td>3.12</td>
<td>2.62</td>
</tr>
<tr>
<td>Length &amp; breadth, Å</td>
<td>40.59 &amp;</td>
<td>40.01 &amp;</td>
<td>37.74 &amp; 24.02</td>
</tr>
<tr>
<td></td>
<td>18.61</td>
<td>22.63</td>
<td></td>
</tr>
<tr>
<td>Length/breadth ratio</td>
<td>0.33</td>
<td>0.28</td>
<td>0.56</td>
</tr>
<tr>
<td>Secondary structure classification(^{c})</td>
<td>alpha</td>
<td>alpha</td>
<td>Alpha</td>
</tr>
<tr>
<td>% polar atoms in interface</td>
<td>48.8</td>
<td>43.0</td>
<td>42.8</td>
</tr>
<tr>
<td>% non-polar atoms in interface</td>
<td>52.2</td>
<td>57.0</td>
<td>57.2</td>
</tr>
<tr>
<td>Segmentation(^{d})</td>
<td>6 (4)(^{e})</td>
<td>5(4)(^{e})</td>
<td>4 (3)(^{e})</td>
</tr>
</tbody>
</table>

\(^{a}\) Computations were made using the A & P monomers of CbnR which constitutes its tetrameric interface. CbnR has a tetramer per asymmetric unit. The DNA binding domain and helix linker (residues 1-89) of CbnR was excluded from calculations.

\(^{b}\) Planarity is a measure of how far the interface residues deviate from a best fit plane. This plane is calculated through the 3-dimensional co-ordinates of the atoms in the interface using principal component analyses.

\(^{c}\) Secondary structure classification represents the secondary structure feature that occurs with the greatest frequency out of alpha, and beta secondary structures in the interface residues.

\(^{d}\) Segmentation is the number of discontinuous segments of the polypeptide involved in the interface interaction.

\(^{e}\) The number in parentheses refers to the number of segments that result when segments containing single or double residues are not considered separate segments.
Dimers interact with one another through helices αH6 and H5, strand β4 and random coil loops. Alpha helix αH6 interacts with its symmetry related partner through hydrogen bonds between the side chains of residues Asn209 and Ser212 in an antiparallel fashion (Table 5.3. and figure 5.3A). The average distance between the backbones of both helices is about 7.5 Å. There were approximately ten direct hydrogen bonds between dimers for both structures (Table 5.3). Although the alpha helix was seen as the major secondary structure interaction, some of the interacting surfaces were also at the N-terminal end of beta strand β4 comprising residues 155-161 and the loop before the beta strand involving residues 147-152. Ser150 (in the loop between β3 and β4) from one dimer interacts with Asn185 (H5), His214 (in the loop between αH6 and β6) and Asp262 (in the loop between β8 and β9) from the other dimer to form an ion-pair network (Table 5.3; Figure 5.3B). A salt bridge is formed between Arg156 (in β4) and Asp262 (in the loop between β8 and β9). A hydrogen bond is formed between Lys148 (in the loop between β3 and β4) and Asp213 (in αH6) (figure 5.3C and Table 5.3). Apart from these direct interactions between residues, there were also many hydrogen bonds mediated by water molecules.

While the interfaces are similar between the two BenM-EBD crystal forms, the fact that some residues are closer or farther away in one structure with respect to the other suggests that there is some flexibility in the interface. Whether the specific structural differences seen here correlate with different conformational states that may be biologically significant is not clear. However, this conformational flexibility may be critically important in allowing LTTRs to change global conformations when their effector ligands are bound. Some of the residues, like Lys148, are involved in the
FIG. 5.3. Interactions at the tetrameric interface of crystal structures A and B. One monomer from a dimer is shown with the backbone and carbon atoms colored purple with the other two-fold related interacting monomer colored yellow. Interactions between residues on the helices of interacting dimers (αH6) are shown in panel A. Hydrogen bonds between Ser150 from one dimer and three other residues from another dimer are shown in panel B, while salt bridges at the tetrameric interface are shown in panel C. Distances between the residues are listed in Table 5.3.
<table>
<thead>
<tr>
<th>Interaction</th>
<th>Distance (Å)</th>
<th>Interaction</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys148-Asp264(OD2)</td>
<td>3.06</td>
<td>Lys148-Asp264(OD2)</td>
<td>4.34</td>
</tr>
<tr>
<td>Lys148(O)-Gly215(N)</td>
<td>3.25</td>
<td>Lys148(O)-Gly215(N)</td>
<td>4.08</td>
</tr>
<tr>
<td>Lys148-Asp213(O)</td>
<td>3.89</td>
<td>Lys148-Asp213(O)</td>
<td>2.90</td>
</tr>
<tr>
<td>Ser150(OG)-Asp262(OD1)</td>
<td>2.55</td>
<td>Ser150(OG)-Asp262(OD1)</td>
<td>2.83</td>
</tr>
<tr>
<td>Ser150(OG)-Asn185(N)</td>
<td>3.10</td>
<td>Ser150(OG)-ASN185(N)</td>
<td>3.11</td>
</tr>
<tr>
<td>Ser150(OG)-Leu184(N)</td>
<td>3.42</td>
<td>Ser150(OG)-Leu184(N)</td>
<td>3.28</td>
</tr>
<tr>
<td>Ser150(N)-His214(O)</td>
<td>3.13</td>
<td>Ser150(N)-His214(O)b</td>
<td>4.59</td>
</tr>
<tr>
<td>Arg156(NH2)-Asp264(OD1)b</td>
<td>4.64</td>
<td>Arg156(NH2)-Asp264(OD2)</td>
<td>2.84</td>
</tr>
<tr>
<td>Asn209 (OD1)-Ser212 (OG)</td>
<td>2.62</td>
<td>Asn209 (OD1)-Ser212 (OG)b</td>
<td>4.05</td>
</tr>
</tbody>
</table>

*a*There are also symmetry related interactions for each bond indicated  

*b*This is not a hydrogen bond but simply a measure of the distance between the atoms in order to compare its equivalent in the counterpart structure
interfaces of both structures, but they pair with different ionic partners. This flexibility may confer a thermodynamic balance between conformations such that effector binding can cause a large structural change while keeping the dimers associated. Interestingly, these same tetrameric interactions were also observed in the modelled full-length structure of DntR, another LTTR (33). The interface parameters such as the interface accessible surface area, planarity, interface residue segments and secondary structure classification were similar to those of the BenM and CbnR structures (Table 5.3). The same helix involved in the tetrameric interactions of CbnR was identical to that of BenM. The predominant segment of residues in the accessible surface interface of DntR, matched residues Gly205 to Lys219 and Lys200 to Glu217 respectively of BenM. Similar structural features were also noticed in the tetrameric interface crystal lattice of the OxyR structure (6). Overall, the tetrameric interface in DntR, OxyR and BenM were comparable to CbnR as they employed similar secondary structural elements and analogous residues.

**Crystal packing and oligomerization**

By applying the crystallographic symmetry operations in both space groups, an extended oligomerization scheme was observed. The structure with the C2221 space group (structure B), which contained two BenM-EBD subunits in the asymmetric unit, has four dimers in the unit cell as shown in figure 5.2B. The other three dimers in the unit cell were built by the following symmetry operations: -x+1,y,1/2-z; -x+1,-y+1,½-z and x,-y+1,-z+I. Although crystal structure A had one subunit in the asymmetric unit, a BenM-EBD dimer could be generated by applying the appropriate crystallographic
symmetry transformations \((-x,y,-z)\). Thus, allowing the monomer to interact with another mobile by an anti-parallel, side by side alignment, making it identical to the dimer in the asymmetric unit of crystal structure B dimer (Figure 5.2B). The other dimers as seen in the complete unit cell of structure A were built by applying the following symmetry operations to the generated BenM-EBD dimer of structure A: \(y,x,1/4-z\); \(y,-y+1, ½-z\); \(-y+1,x,3/4+z\) and \(x,y,z+1\). The tetrameric interface and interactions seen in structure A utilized similar surface areas and amino acid residues as structure B (Figure 5.2 and Table 5.2).

The formation of a continuous oligomer using the tetramerization interface was visualized by expanding the asymmetric units to the full unit cells of both crystal structures. A high-order oligomeric array could be computed for both crystal structures as shown in their unit cells by applying appropriate crystallographic symmetry operators (figure 5.4). The DNA binding-domains of the different subunits of CbnR do not show any interaction with each other across the tetrameric interface (25). Further, modelling of CbnR tetramers into the structure A oligomeric array does not introduce any steric clashes from the DNA-binding domain with neighboring tetramers. Interestingly, the CbnR crystal structure does not utilize the tetramerization interface as crystal contacts. This may account for the success in crystallizing this particular full-length LTTR.

Figure 5.5 shows a schematic of two possible pathways by which extended oligomerization could occur for BenM and LTTRs in general. In pathway A, the DNA binding domains interlock two dimers to stabilize the tetramer as a single unit. Two tetramerization sites are left accessible to other tetramers. So at high protein concentrations, linear arrays would form that could ultimately precipitate. The array
FIG. 5.4. Unit cell representations of BenM-EBD structures A (panel A) and B (panel B). The unit cell is outlined in black and a, b and c crystallographic cell axes are shown. Structure A is colored by monomers as it has a monomer in the asymmetric unit. Structure B is colored by dimers as it has a dimer in the asymmetric unit.
**FIG. 5.5.** Proposed schematic for the oligomerization of BenM in light of the CbnR structure (1IZ1). Two avenues by which oligomerization could take place are shown. Panel A shows the high-order oligomer formation in which there are no interactions of the DNA binding domains (represented by the open and closed ovals) in the dimer-dimer interface (tetramerization interface). Two oligomerization interfaces are present, one interface within a tetramer and a second between two tetramers. Panel B shows an alternate route for oligomer formation, but this time the DNA binding domains interact across all of the tetramerization interfaces. The tetramers are colored red, yellow, green and blue. The DNA-binding domains assume two conformations in the structure of CbnR, and thus are represented as two different ovals. The monomer-monomer interface (the dimer interface in this text) is shown as a straight line, while the tetramerization interface of each monomer is curved.
could dissociate readily to individual tetramers. The DNA-binding domains assume two different conformations in the structure of CbnR (25). In pathway B, domain swapping can occur such that the DNA-binding domains now interact with dimers from nearby molecules, but in such a way that an individual tetramer is no longer formed. In this situation, once formed, the array would not likely dissociate as readily as the array in pathway A. We have not evaluated whether pathway B is conformationally possible. A permutation of pathway B not shown would include the formation of conformationally identical DNA-binding domains (both ovals open and both ovals filled).

CONCLUSION

Protein molecules tend to randomly form crystal-packing contacts during crystallization (4), and so in most cases, little biological significance is assigned to these contacts. However, our analysis presented here used the existing knowledge of the biological function of this family of proteins to evaluate the significance of these interactions in the context of the fundamental properties that have been shown to characterize physiological protein-protein interactions (18, 19).

Refinement of two BenM-EBD structures of different space groups to a resolution of 2.2 Å and 2.7 Å from relatively high pH crystallization conditions has allowed us to provide a plausible description of the tetrameric interface in this LysR-type transcriptional regulator. Identification of residues in this interface and interactions that lead to the formation of a tetrameric molecule have been established. It has also been shown that the crystal packing observed in both structures uses this same tetrameric interface for the formation of high-order oligomers as observed in the crystal lattice of
both structures. We suggest that this oligomerization scheme is found in LTTRs as a class and may be the cause of the insolubility problems associated with BenM and probably other proteins in this family. An LTTR tetramer leaves two additional tetramerization sites exposed, which may lead to further oligomerization. In the tetrameric CbnR structure (25), the interactions in the tetrameric interface employ significantly similar surface contacts as BenM-EBD, thus substantiating the biological significance of this interaction as seen in the structures of the truncated effector binding domain of BenM presented here.

Several general observations can be made concerning biological relevance applicable to LTTRs. First, while the monomer-monomer (dimer interactions) and the dimer-dimer (tetramer interactions) are structurally well conserved in BenM and CbnR, the sequence similarity between these regions is low and thus the local nature of the interactions are different between the LTTRs. Monomer-monomer contacts have higher sequence similarity than dimer-dimer contacts. These differences are desirable from a biological standpoint as an organism would not want mixed dimers/tetramers of LTTRs that could produce deleterious metabolic consequences. High order oligomerization could play a role in transcriptional regulation if it takes place in a cell, though no one has observed this. Oligomerization beyond tetramers is unlikely because LTTRs negatively regulate their own transcription. Thus, once enough monomer is made to organize into a biologically active tetramer, little further transcription of the LTTR’s gene should take place, which would limit the concentration of protein that could create higher-order oligomers. While oligomerization may not be an issue in cells; it becomes interesting to the crystallographer as it creates problems with solubility of the protein at the near
millimolar concentrations routinely used for crystallization trials. This problem is not only manifested in the full-length proteins where it is extreme, but even in the truncated effector binding domains. Mutagenesis at the tetramerization interface is not a fully acceptable solution as changes that prevent oligomer formation are likely to interfere with biological function and would be structurally irrelevant. We are attempting to cap the exposed tetramer interfaces using antibody fragments as one approach to solving the problem. Alternatively, crystallization attempts might be made at submillimolar concentrations in the presence of interface destabilizing buffer components. We anticipate that structural studies of LTTRs will be hampered as long as the oligomerization issue in this family is not fully addressed.
ACKNOWLEDGEMENTS

We are grateful to the staff at the Hauptman-Woodward Institute for performing the high-throughput crystallization screens and the staff at the Structural Biology Centre Collaborative Access Team (SBC-CAT) 19-BM beamline (Advanced Photon Source, Argonne National Laboratory) for assistance in X-ray data collection and the use of their facility. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, and Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38. We also thank Michelle Momany for assistance in collecting X-ray data and Peter Horanyi for useful suggestions concerning the manuscript preparation. This research was supported by grants MCB 0346422 to CM and MCB 0516914 to ELN from the National Science Foundation.
REFERENCES


Transcriptional cross-regulation of the catechol and protocatechuate branches of 
the beta-ketoadipate pathway contributes to carbon source-dependent expression 

transcriptional activation by one regulatory protein in response to two 

Protein Sci. 6:2261-3.

5:1269-74.


effector-binding domains of BenM and CatM, LysR-type transcriptional 


CHAPTER 6

DISSERTATION SUMMARY

The goal of this dissertation was to establish the structural and functional differences between two LysR-type transcriptional proteins, BenM and CatM. As described in chapter 2, the functional differences between both proteins were analysed at the benABCDE promoter. CatM was shown to activate the benABCDE promoter under certain mutational conditions that allowed growth on benzoate in the absence of BenM. Two of these mutations were in catM that generated variants, CatM(V158M) and CatM(R156H). These substitutions affected regulation differently and demonstrated the importance of balanced expression from multiple promoters in the degradation of benzoate. In light of these studies, a model of CatM-regulated benA expression was proposed.

In chapter 3, structures of the BenM- and CatM-effector binding domains were solved with and without muconate and/or benzoate bound to the protein. Structural analyses highlighted the importance of conformational changes in response to these effectors in modulating transcription. The muconate binding site of both BenM and CatM was established. An additional benzoate site was found in BenM but not CatM. From these studies, a model was proposed for the synergistic transcriptional activation by BenM in response to muconate and benzoate. The studies also provided a structural basis for why BenM but not CatM responds to benzoate.

These structural studies of the effector binding domains of BenM and CatM further served as a basis for analyzing the altered regulatory roles of three variants of CatM and BenM. The crystal structures of CatM(V158M), CatM(R156H) and
BenM(R156H/T157S) effector binding domains were determined and compared to the wild-type structures. The locations of these substitutions were shown to serve as important anchor points of the protein in inducing conformational changes that alter their regulatory roles.

In light of the structural studies, the rationale behind the insolubility problems encountered by LysR-type regulatory proteins was established in chapter 5. The elucidation of two BenM-EBD structures generated from high pH conditions established the tetrameric interface of BenM and further showed how these proteins could form high-order oligomers that lead to insolubility problems in BenM. Collectively, the studies described in this dissertation clarify the complex and overlapping roles of two important LysR-type paralogs in *Acinetobacter baylyi* ADP1.
List of crystal structures submitted to the Protein data bank (PDB)

<table>
<thead>
<tr>
<th>Structures</th>
<th>PDB ID code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CatM-EBD</td>
<td>2F7B</td>
</tr>
<tr>
<td>2 CatM-EBD-\textit{cis,cis-}muconate</td>
<td>2F7C</td>
</tr>
<tr>
<td>3 BenM-EBD-SeMet</td>
<td>2F6P</td>
</tr>
<tr>
<td>4 BenM-EBD</td>
<td>2F6G</td>
</tr>
<tr>
<td>5 BenM-EBD-benzoate</td>
<td>2F78</td>
</tr>
<tr>
<td>6 BenM-EBD-\textit{cis,cis-}muconate\textsuperscript{a}</td>
<td>2F7A</td>
</tr>
<tr>
<td>7 BenM-EBD-high pH\textsubscript{1} (structure A)</td>
<td>2F97</td>
</tr>
<tr>
<td>8 BenM-EBD-high pH\textsubscript{2} (structure B)</td>
<td>2F8D</td>
</tr>
<tr>
<td>9 CatM(V158M)-EBD</td>
<td>2H98</td>
</tr>
<tr>
<td>10 BenM(R156H/T157S)-EBD (structure A)</td>
<td>2H99</td>
</tr>
<tr>
<td>11 BenM(R156H/T157S)-EBD (structure B)</td>
<td>2H9B</td>
</tr>
<tr>
<td>12 CatM(R156H)-EBD-\textit{cis,cis-}muconate\textsuperscript{a}</td>
<td>2H9Q</td>
</tr>
</tbody>
</table>