GENE EXPRESSION AND DOSAGE: DISTINCT MECHANISMS REGULATE BENZOATE DEGRADATION IN *ACINETOBACTER BAYLYI* ADP1

by

SARAH HOPE CRAVEN

(Under the Direction of Ellen L. Neidle)

ABSTRACT

Multiple mechanisms regulate aromatic compound degradation in the soil bacterium *Acinetobacter baylyi* ADP1. Herein, complementary mechanisms controlling the consumption of benzoate are explored: 1) transcriptional regulation and 2) chromosomal amplification of catabolic genes. The former highlights BenM, a member of the LysR family of regulators controlling transcription of a complex regulon for benzoate degradation in ADP1. BenM activates gene expression in response to two effectors, benzoate and muconate. These effectors act synergistically to allow high-level gene expression, a feature not characterized in any other member of this broad family. Here, the unique effector profile of BenM is explored. Genetic studies, combined with structural work, elucidate the role of two distinct effector-binding sites in BenM. Targeted mutagenesis confirms the physiological relevance of a hydrophobic benzoate binding pocket and pinpoints critical residues, Arg160 and Tyr293, essential for benzoate-dependent gene expression and synergistic transcriptional activation. BenM variants that activate transcription in the absence of inducer are examined: BenM(R156H), BenM(R225H), and BenM(E226K). Structural studies of BenM variants reveal subtle conformational changes associated with transcriptional activation.

The recent discovery of gene amplification in ADP1 highlights an alternate strategy of genetic regulation. In strains lacking wild-type regulation, chromosomal amplification of catabolic genes can compensate for low transcript levels. Here, a wide spectrum of amplification events were studied using multiple approaches uniquely available in the tractable ADP1 model system. Approximately 50 amplification mutants with enhanced benzoate catabolism were characterized. Each harbors multiple copies of a genomic region, ranging from 12 kbp to more than 300 kbp, encompassing the *ben* and *cat* functional genes. Analysis of the precise sequence at amplicon endpoints identifies the DNA sequence involved in formation of the original duplication. In all cases examined, these sequences indicate genetic duplication was initiated by a homology-independent illegitimate recombination mechanism. Recurrence of exact duplication junctions in multiple mutants reveals site preference in duplication formation, a mechanism we designate position-specific illegitimate recombination. The novel experimental tools developed here will facilitate a rapid and thorough examination of the full spectrum of gene amplification events contributing to ADP1 genome plasticity, as well as enable studies into the relevant recombination mechanisms.

INDEX WORDS: Gene regulation, LysR-type transcriptional regulator, *Acinetobacter*, Gene amplification, Genome plasticity, Aromatic compound catabolism

GENE EXPRESSION AND DOSAGE: DISTINCT MECHANISMS REGULATE BENZOATE DEGRADATION IN *ACINETOBACTER BAYLYI* ADP1

by

SARAH HOPE CRAVEN

B.S., High Point University, 2004

B.S., High Point University, 2004

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

2009

GENE EXPRESSION AND DOSAGE: DISTINCT MECHANISMS REGULATE BENZOATE DEGRADATION IN *ACINETOBACTER BAYLYI* ADP1

by

SARAH HOPE CRAVEN

Major Professor: Ellen Neidle

Committee: Timothy Hoover
Michael McEachern
Cory Momany
Anne Summers

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2009

ACKNOWLEDGEMENTS

I gratefully acknowledge the members of my doctoral committee: Cory Momany, Anne Summers, Tim Hoover and Mike McEachern. I appreciate the advice and guidance each provided throughout my research. In particular, I express gratitude to my thesis advisor, Ellen Neidle. I am grateful to Ellen for all she has done to promote my scientific and personal growth. I thank current and former members of the Neidle lab: Obidimma Ezezika, Katie Pennington, Sandra Haddad, Laura Cuff and K.T. Elliott. Thank you for teaching me, for learning alongside me, and for commiseration and encouragement. I acknowledge Vera Patel, Jenny Taylor, Ruth Hall, and Jen Hiras—students who are not only great scientists, but will remain my true friends.

I would like to acknowledge Kelli Sapp and Charles Warde—two special professors who continue to guide my professional and personal growth. They have set the example, as mentors and as scientists, to which I aspire. The admiration and respect I hold for them continues to have a tremendous impact on my development as a scientist.

Most importantly, I would like to acknowledge the unwavering support of my family. It is only with the love and support from my parents, my grandparents, and my brother and sister, that I have achieved this goal. To my parents—thank you for your constant patience and love. You amaze me each day with your kindness and faith. No matter where the future takes me, my heart will always be home in North Carolina. Finally, I offer my love and thanks to Daniel. Thank you for making me laugh amid tears and for encouraging me when the end seemed out of reach. Each day is more fun with you in my life. I am blessed to be surrounded by people who make me a better a person. We made it!

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

*I gratefully acknowledge Dr. Cory Momany for structural figures included throughout this dissertation.

CHAPTER 1

INTRODUCTION

The soil bacterium *Acinetobacter baylyi* ADP1 thrives on a wide range of organic acids and aromatics—plant-derived compounds typically found in abundance in soil environments. The nutritional versatility of this organism is apparent at the genome level, with more than 20% (>650 genes) of its 3.5 Mbp genome dedicated to catabolism and another 10% (~320 genes) attributed to transport. The streamlined genome of ADP1, <60% the length of aerobic *Pseudomonas* counterparts with similar nutritional spectra, makes this organism unusually well-adapted to a dynamic and competitive soil environment.

With this catabolic versatility, however, comes a need for strict regulation of gene expression and enzyme production. The importance of control is two-fold: 1) the cost of synthesizing enzymes required for catabolism is taxing and only need be invested when the substrate is present and 2) unbalanced expression of catabolic pathways can lead to the accumulation of toxic intermediates within the cell. Thus, organisms with efficient regulation of gene expression likely hold a competitive advantage in a dynamic environment. Myriad sophisticated regulatory networks control catabolism in *Acinetobacter*—including carbon catabolite repression to ensure hierarchical consumption of mixed carbon sources (Dal *et al.*, 2002), synergistic gene expression in response to multiple effector compounds (Bundy *et al*., 2002), cross-regulation of multiple operons by a single metabolite (Siehler *et al.*, 2007), and increases in dosage of catabolic genes via genetic amplification (Reams and Neidle, 2003).

This dissertation elaborates on two distinct, yet complementary types of genetic regulation in ADP1: 1) control mechanisms of LysR-type transcriptional regulators (LTTRs) and 2) the chromosomal amplification of catabolic genes. Both involve control of a 20 kbp supraoperonic cluster of genes required for benzoate consumption, the *ben* (benzoate) and *cat* (catechol) genes. The former details a classic type of bacterial regulation by DNA binding transcriptional regulators, BenM and CatM. The latter details a transient regulatory phenomenon, allowing enhanced benzoate catabolism by an increase in gene dosage of the *ben* and *cat* functional genes. The following chapters describe these disparate, yet interrelated topics and outline novel developments in both.

Regulatory Studies in ADP1: Overview and Organization

For ease of interpretation, the following dissertation is divided by topic: Chapters 2 and 3 describe mechanisms of LTTR-mediated regulation while Chapters 4 and 5 describe regulation by genetic amplification. Specifically, the second chapter provides a review of LTTRs, a diverse and widespread family of bacterial regulators. The chapter emphasizes the importance of this type of regulation in *Acinetobacter* and details the 44 predicted LTTRs encoded by the ADP1 genome, including predictions of their regulatory function and target promoters. This chapter further provides a review of the BenM-CatM regulon in ADP1 and thus provides the basis for Chapter 3—a genetic and structural study of effector binding by the BenM regulator.

Chapter 4 and Chapter 5 focus on consequences of genetic duplication and amplification in bacteria. Gene amplification, a widespread phenomenon in bacteria and higher organisms, was recently discovered in ADP1 and affords enhanced benzoate consumption in strains lacking the BenM and CatM regulators. By way of introduction, Chapter 4 highlights diverse examples of

gene amplification in bacteria, with an emphasis on clinical manifestations of gene amplification in pathogenic microorganisms. Despite the widespread importance of genetic duplication and amplification, many questions about the frequency, distribution and mechanisms underlying these events remain unclear. Thus, Chapter 5 of this dissertation describes continued development of strain ADP1 as a tractable model for studies of gene duplication and amplification. The final chapter offers a brief summary and asserts some broader implications of the work included herein.

CHAPTER 2

LYSR HOMOLOGS IN *ACINETOBACTER*: INSIGHTS INTO A DIVERSE FAMILY OF TRANSCRIPTIONAL REGULATORS[1]

[1] Craven, S.H., O.C. Ezezika, C. Momany, and E.L. Neidle. 2008. In: *Acinetobacter* Molecular Biology. Caister Academic Press, pp. 163-202. Reprinted here with permission of publisher.

Overview and General Features of LysR-type Regulation

LysR-type transcriptional regulators (LTTRs) were first recognized as a family of homologous proteins nearly twenty years ago (Henikoff *et al.*, 1988). As the characterization of regulatory protein sequences increased, it soon became evident that members of the LysR family are widely distributed in prokaryotes and capable of regulating a variety of metabolic functions (Schell, 1993). While these regulators share common features, new studies continue to reveal distinctive aspects of individual LTTRs and their regulons (Lee *et al.*, 2006, Peck *et al.*, 2006, Smart & Bauer, 2006, Xiao *et al.*, 2006, Yang *et al.*, 2006). With bioinformatic approaches used to assess the complete genomes of diverse organisms, the importance of LysR-type regulation is further underscored by evidence that this family represents the most common type of bacterial transcriptional regulator.

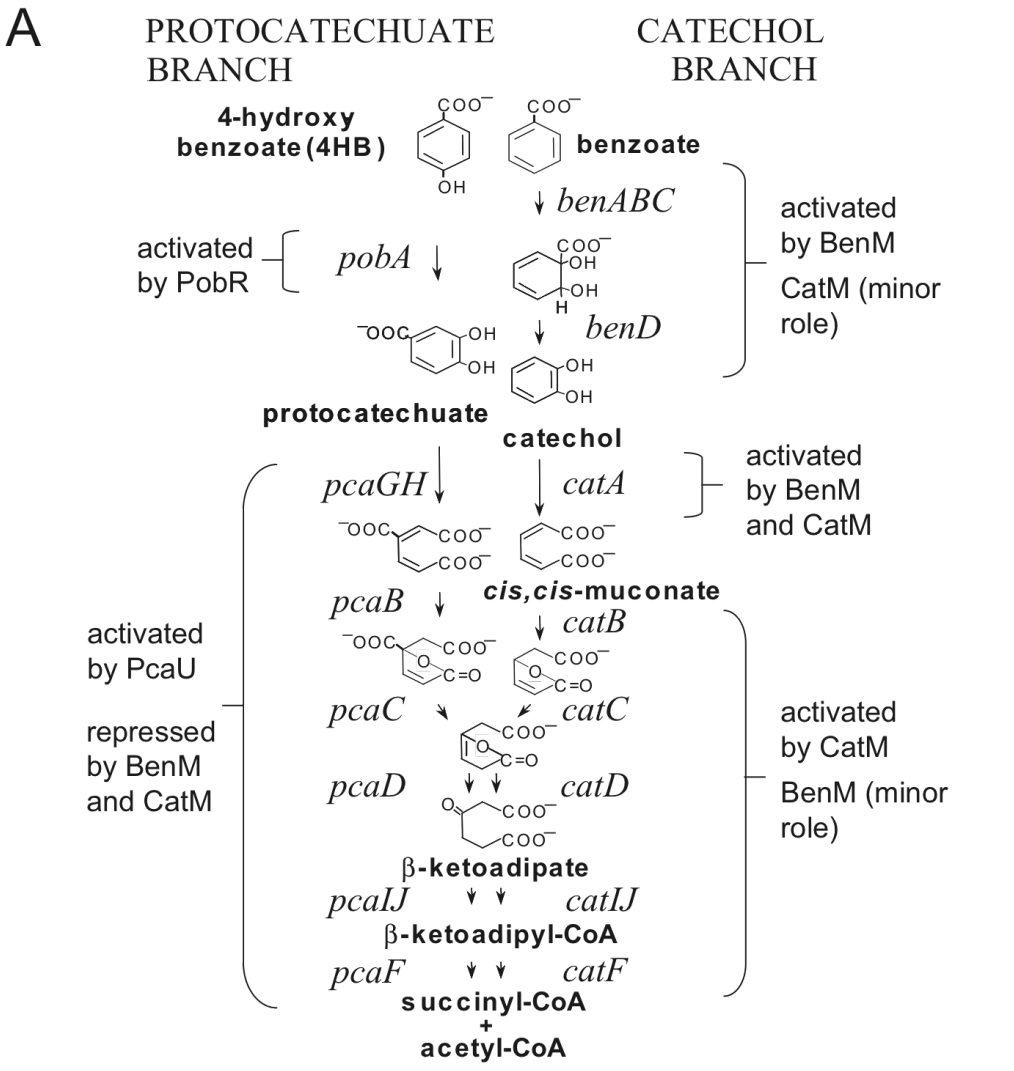
The 26,046 archaeal and bacterial regulatory proteins in the UniProt database were classified into 16 regulatory families: AraC/XylS, ArsR, AsnC, Cold shock domain, CRP-FNR, DeoR, GntR, IclR, LacI, LuxR, LysR, MarR, MerR, NtrC/Fis, OmpR and TetR (Pareja *et al.*, 2006). This analysis, which includes 230 prokaryotic genomes, identified nearly 3900 LysR-type regulators, representing 15% of the identified regulatory sequences. While LTTRs are not common in the archaea, they are prevalent in the proteobacteria. Strains of *Agrobacterium*, *Burkholderia, Escherichia, Pseudomonas,* and *Sinorhizobium* each have genomes predicted to encode approximately 40 to 120 family members. In many bacteria, LTTRs comprise approximately 20-30% of the regulators encoded by a single genome.

In the soil bacterium *Acinetobacter baylyi* ADP1, the overall repertoire of transcriptional regulators is estimated at 174 proteins, including 44 predicted LTTRs (Barbe *et al.*, 2004, Pareja *et al*., 2006, Vaneechoutte *et al.*, 2006). In *A. baylyi* ADP1, the next most abundant type of

transcriptional regulator is the AraC/XylS activator, with 31 putative family members (Gallegos *et al.*, 1993, Martinez-Bueno *et al.*, 2004, Pareja *et al*., 2006). The other regulatory families are each predicted to include 0 to 19 members in ADP1. Although the genomes of additional *Acinetobacter* strains are not yet available, the dependence on LysR-type regulators should extend throughout the genus. Therefore, it was not surprising that a *lysR*-type gene was discovered in the 86-kb drug resistance island of *A. baummannii* AYE (Fournier *et al.*, 2006). In this book chapter we explore the possible roles of LTTRs by examining the sequence and context of their genes in the ADP1 chromosome.

A. baylyi ADP1 is a good model organism for assessing LTTR function, and one of its regulators, CatM, was among the first in the family to be studied (Neidle *et al.*, 1989). The ability to discern whether CatM is an activator or repressor was initially obscured by the presence of BenM, a paralog with overlapping function (Collier *et al.*, 1998, Romero-Arroyo *et al.*, 1995). It is now clear that the LTTRs are dual-action regulators that can repress and activate transcription. BenM and CatM activate genes for the degradation of benzoate and related aromatic compounds in response to the appropriate effectors (Figure 2.1). Additionally, they are negatively autoregulated and contribute to the repression of genes used to degrade alternative aromatic carbon sources (Brzostowicz *et al.*, 2003). Here we review the BenM-CatM regulon to illustrate an elaborate and elegant regulatory circuit.

Recent structural studies of the effector-binding domains (EBDs) of BenM and CatM are also described in this chapter (Ezezika *et al.*, 2007a). Despite the importance of LTTRs, very few have been structurally characterized. To date, there is a single known structure for a full-length LTTR, CbnR (Muraoka *et al.*, 2003). Additionally, there are several structures known for


A
PROTOCATECHUATE BRANCH
CATECHOL BRANCH
4-hydroxy benzoate (4HB)
benzoate
benABC
activated by PobR
pobA
activated by BenM
CatM (minor role)
benD
protocatechuate
catechol
pcaGH
catA
activated by BenM and CatM
cis,cis-muconate
pcaB
catB
activated by PcaU
repressed by BenM and CatM
pcaC
catC
pcaD
catD
activated by CatM
BenM (minor role)
β-ketoadipate
pcaIJ
catIJ
β-ketoadipyl-CoA
pcaF
catF
succinyl-CoA
+
acetyl-CoA

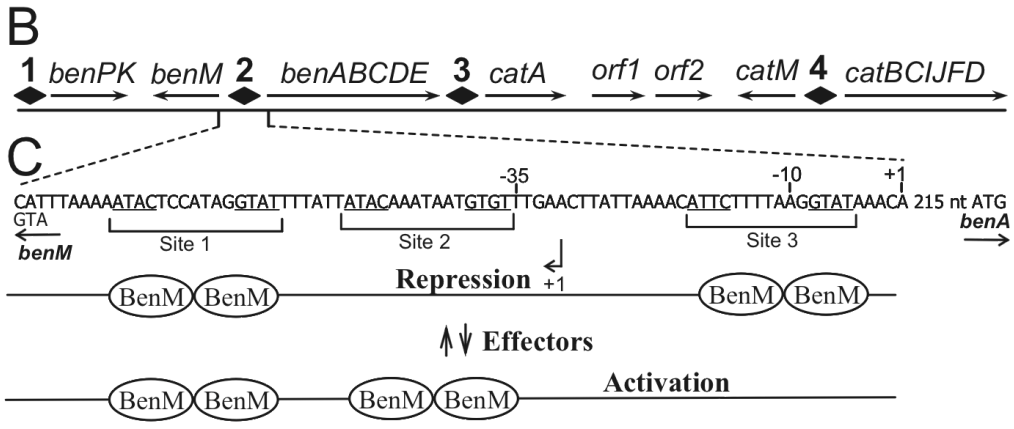

B
1 benPK benM 2 benABCDE 3 catA orf1 orf2 catM 4 catBCIJFD
C
-35
-10
+1
CATTTAAAAATACTCCATAGGTATTTTATTATACAAATAATGTGTTTGAACTTATTAAAACATTCTTTTAAGGTATAAACA 215 nt ATG
GTA
benM
Site 1
Site 2
Site 3
benA
Repression
+1
BenM
Effectors
Activation

Figure 2.1: The BenM-CatM regulon in *A. baylyi* ADP1. (A) Degradation proceeds via two branches of the β-ketoadipate pathway. Regulatory proteins that control gene expression are indicated. Enzymes encoded by the *ben* and *cat* genes are organized in an approximately 20 kbp chromosomal cluster **(B, not drawn to scale)**. BenM and CatM regulate four promoters (diamonds 1-4). The *benPK* operon encodes a porin and a permease of the major facilitator superfamily, and *benE* encodes a membrane protein of unknown function (Clark *et al*., 2002; Collier *et al*., 1997). At promoter regions 1 and 3, the loss of CatM or BenM can be functionally compensated by the presence of the other regulator. At promoter regions 2 and 4, either BenM or CatM, respectively, is the major regulator, and the second regulator plays a minor role in gene expression. Neither CatM-regulated *benABCDE* expression nor BenM-regulated *catBCIJFD* expression allows benzoate to be consumed at wild-type growth rates. **(C)** In the intergenic *benMA* region, there are three potential binding sites for BenM and CatM. Site 1 matches a consensus sequence (underlined) for LysR-type regulators. The sequences in Sites 2 and 3 differ by a single nucleotide that reduces the extent of dyad symmetry. Above the DNA sequence, the transcription initiation site (+1) and the promoter regions (-10,-35) are shown for *benA*. The initiation site for the divergently transcribed *benM* is also indicated (+1 below an arrow). The model for BenM binding that causes *benA* repression in the absence of effectors and *benA* activation in the presence of effectors is based on studies discussed in the text (Bundy *et al*., 2002; Collier *et al*., 1998; Ezezika *et al*., 2006).

the regulatory domains of LTTRs (Choi *et al.*, 2001, Smirnova *et al.*, 2004, Verschueren *et al.*, 1999, Tyrrell *et al.*, 1997). The tendency of LTTRs to aggregate in high concentration solutions can cause problems that preclude structural determination by established X-ray crystallographic methods. Moreover, approaches used to circumvent such problems have included crystallization conditions that prevent the proteins from binding to their natural inducers. Most LTTRs attain their active conformation by binding small-molecule effectors, such as a metabolite to which CatM and BenM bind, *cis,cis*-muconate (hereafter called muconate). Our structural studies of the CatM- and BenM-EBDs are the first to characterize the effector-binding site of any LTTR bound to its natural inducer. Here we review the conformational changes that occur upon ligand binding and demonstrate how such changes can propagate to affect the global conformation of an active tetramer.

Another novel aspect of our structural investigation concerns the ability of BenM, a single regulatory protein, to activate transcription synergistically in response to two effectors, benzoate and muconate (Bundy *et al.*, 2002). Surprisingly, two distinct binding sites were discovered in BenM-EBD that can explain the basis of the synergism (Ezezika *et al*., 2007a). While many regulatory proteins respond to multiple effectors, the prevailing paradigm is that a regulator interacts with a single compound at one time. However, this type of transcriptional synergism provides the cell with a rapid mechanism for integrating cellular signals, and it may be more widespread than currently known.

The BenM-CatM regulon serves to highlight distinctive and general features of LysR-type regulation. To convey a broader perspective, we also comment on interesting regulatory studies from other organisms. However, this chapter is not intended to provide a comprehensive review of LTTRs. Rather, the focus on *Acinetobacter* regulation should complement previous

publications that discuss this important regulatory family (Gerischer, 2002, Schell, 1993, Tropel & van der Meer, 2004).

LTTR proteins

Most LTTRs contain 280 to 350 amino acids and have the greatest level of sequence conservation in the N-terminal DNA-binding domain. The structural characterization of this domain in one full-length LTTR, CbnR, confirmed the presence of a winged-helix-turn-helix (wHTH) motif (Muraoka *et al*., 2003). This motif is one of several variations on an open tri-helical bundle that forms the core of the most commonly used DNA-binding structural scaffold (Aravind *et al.*, 2005, Brennan, 1993). The name of the HTH motif refers to the second and third helices and the sharp bend between them. Interestingly, while the turn region is made up of only a few residues, it appears to play a critical role in activating transcription via direct contact with RNA polymerase (Jourdan & Stauffer, 1998, Lochowska *et al.*, 2004).

The third helix of the bundle, known as the recognition helix, typically inserts into the major groove of the DNA. This helix is followed by the region designated the wing, a beta-strand hairpin connected by outstretched loops. The role of the wing in DNA recognition and binding remains unclear for LTTRs. However, wing-mediated interactions with the minor groove have been demonstrated to help position some proteins on DNA (Huffman & Brennan, 2002). Protein-DNA interactions for LTTRs most likely involve residues in the first helix of the bundle as well as those in the wHTH motif (Muraoka *et al*., 2003, Kullik *et al.*, 1995, Lochowska *et al.*, 2001). In its entirety, the DNA-binding domain of an LTTR is made up of approximately 60 amino acids at the N-terminus.

The CbnR structure reveals a long linker helix (residues 59-87) that connects the DNA-binding domain to a regulatory domain, extending to the C-terminus, which is made up of two distinct subdomains (Muraoka *et al*., 2003). The protein fold of this CbnR regulatory domain, which is essentially that of periplasmic binding proteins (Quiocho & Ledvina, 1996), is conserved in all other LTTR regulatory structures that have been characterized to date, CysB, OxyR, DntR, Cbl, CatM and BenM (Choi *et al*., 2001, Smirnova *et al*., 2004, Tyrrell *et al*., 1997, Verschueren *et al*., 1999, Ezezika *et al*., 2007a, Stec *et al.*, 2006). This high level of structural conservation is remarkable because there is great variability among the sequences of this region in different LTTRs. For example, the sequence identity between the regulatory domain of BenM and the comparable regions of CysB and OxyR are only 12 and 21%, respectively. This sequence variation reflects the functional differences among the regulators.

In most LTTRs, the regulatory domain serves to bind small-molecule effectors that induce conformational changes affecting transcription (Schell, 1993). The ability of different LTTRs to bind diverse effectors correlates with differences in the primary sequences of the regulatory domains. Regulators that bind similar compounds can be grouped in LTTR subfamilies based on sequence homology. The structural characterization of the EBDs of BenM and CatM bound to their cognate effectors, benzoate and muconate, confirm the prediction that an effector can bind in the cleft between the two domains of the regulatory region (Ezezika *et al*., 2007a). Such binding is shown for muconate in the BenM-EBD structure depicted in Figure 2.2. This structure is described more completely at the end of this chapter and in Chapter 3 of this thesis.

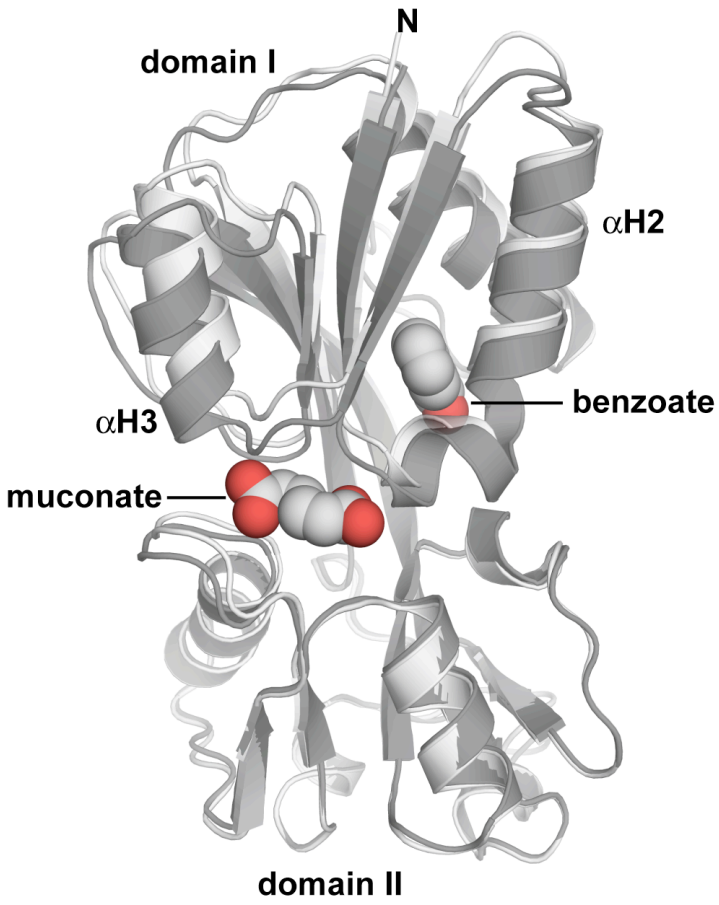


Figure 2.2: Ribbon representation of the α-carbon backbone of two BenM-EBD subunits. One subunit is bound to muconate and benzoate (dark gray) and the other is not bound to effectors (light gray). Domain II of the effector-bound structure was superimposed on that of unbound structure. α-Helices (αH) and 3_{10}-helices (H) that are important for conformation changes associated with effector binding are labeled and numbered. There are nine helices per subunit, numbered consecutively from the N-terminus (Ezezika *et al*., 2007a).

Most LTTRs function as homotetramers or homodimers, and several studies indicate that the C-terminal regions are important for oligomer formation (Kullik *et al*., 1995, Lochowska *et al*., 2001, Schell *et al.*, 1990). Residues at the dimerization interface have been identified from several LTTR structural studies. In most of these X-ray crystallography investigations, the regulatory proteins crystallized as dimers. In the case of the full-length CbnR, the protein crystallized as a tetramer allowing additional characterization of the tetramerization (dimer-dimer) interface. Interestingly, some amino acid substitutions that enable LTTRs to activate transcription in the absence of their effectors map to regions of the dimerization interface (Akakura & Winans, 2002a, Choi *et al*., 2001). Other CysB and OxyR variants that cause constitutive phenotypes appear to bind DNA as dimers in contrast to their tetrameric wild-type counterparts (Kullik *et al*., 1995, Lochowska *et al*., 2001). Therefore, alterations in the oligomeric conformation may facilitate the formation of a protein-DNA complex resembling one that would normally be attained via ligand binding to the LTTR.

The CbnR structure reveals that within the tetramer, there are two different subunit configurations, an extended form and a compact form (Muraoka *et al*., 2003). The tetramer can be viewed as a dimer of a dimer with one subunit in each configuration. The DNA-binding domains do not interact with each other but are held in position by the four centrally located EBDs. Each EBD should contract as it clamps down on the effector between its subdomains, and the additive effect of four such contractions should bring two pairs of DNA-binding domains closer together (Ezezika *et al*., 2007a). This movement could account for the observed changes that occur in LTTR-DNA interactions in response to effector binding. A similar conclusion was drawn from structural studies of DntR (Smirnova *et al*., 2004). In this case, the LTTR crystallized as a full-length protein, but a high-resolution structure of the DNA-binding domain

could not be obtained. Structural modeling suggests that the DNA-binding domains of DntR are closer together than those of CbnR. The DntR structure might resemble an effector-bound LTTR since acetate or thiocyanate was present in the presumed effector-binding site. While these compounds derive from the crystallization conditions, they could mimic the effect of the native inducer.

Genetic organization and promoter interactions

LTTRs, which bind DNA in the presence or absence of their cognate effectors, often occupy multiple binding sites in the operator-promoter regions of their target genes. Effector interactions can elicit alterations in the DNA-binding position(s) of the oligomeric protein or in the number of subunits that are bound. The exact details of these protein-DNA interactions vary for different LTTRs. In many cases they also vary for the same LTTR acting at different promoters. Here, interactions of BenM at the *benA* promoter region of *A. baylyi* ADP1 illustrate the type of protein-DNA interactions that can occur.

The *benM* gene is transcribed divergently from the *benABCDE* operon (Figure 2.1 B). Divergent promoters are common for LTTR genes, and this arrangement coordinates negative autoregulation with the positive control of an adjacent target. DNase I footprint patterns indicate that in the absence of its effectors a BenM tetramer protects two areas of the *benA* region from cleavage, Site 1 and Site 3 in Figure 2.1 (Bundy *et al*., 2002). In this region there are three sequences identical or similar to ATAC-n7-GTAT, and the DNA-binding domain of one BenM subunit appears to recognize half this dyadic sequence. This sequence matches the general LTTR consensus-binding site, T-n11-A within a small stretch of dyad symmetry. Without effectors, BenM represses *benA* expression by blocking the –10 region of its promoter. With effectors, the

protein-DNA complex changes such that BenM protects Sites 1 and 2 from DNase I cleavage. This alteration most likely facilitates the ability of BenM to recruit RNA polymerase to activate *benA* transcription. BenM binding represses *benM* expression in both cases.

For other LTTRs, there are notable variations concerning the sequence, position and number of binding sites in the DNA, as previously discussed (Akakura & Winans, 2002b, Schell, 1993, Tropel & van der Meer, 2004). An important role is often ascribed to DNA bending. DNase I footprints of BenM at the *benA* promoter in the absence of effectors reveal hypersensitive cleavage sites in the intervening sequence between protected Sites 1 and 3 (Bundy *et al*., 2002). This cleavage pattern is consistent with the hypersensitive sites being positioned along the edge of an exposed DNA loop that forms as the tetramer bends the DNA. The effectors shift the position of the regulatory protein on the DNA and eliminate hypersensitive cleavage sites, indicating a relaxation of the DNA.

LTTR-mediated DNA bending usually follows one of two general patterns. The first is exhibited by regulators such as OccR, CatR, and CysB, which can induce a sharp-angle DNA bend whose angle diminishes upon addition of the appropriate effector (Hryniewicz & Kredich, 1995, Parsek *et al.*, 1995, Wang *et al.*, 1992). Conversely, the addition of an effector to some regulators, such as NodD, can increase the DNA bend angle of the LTTR-promoter complex (Chen *et al.*, 2005). In both cases, ligand-induced conformational changes alter promoter architecture to activate transcription.

To clarify the role of DNA bending, mutations that increase or decrease the angle of a DNA bend have been engineered and evaluated for their effect on transcription. In one study, the spacing was altered between two operator sites that anchor a NodD tetramer on DNA controlling *nodA* expression (Chen *et al*., 2005). Some mutations that increased the bend angle in a similar

fashion to the effector, naringenin, enabled *nodA* to be transcribed without an effector. Analysis of several different mutations demonstrated a correlation between the NodD-mediated DNA bend angle and transcriptional activation. In another study, OccR-induced DNA bending was modulated through site-specific mutagenesis of the *occQ* operator-promoter region (Akakura & Winans, 2002b). OccR-binding sequences were altered in ways predicted to lock the protein-DNA complex into a transcriptionally active or inactive conformation containing a low-angle or high-angle DNA bend, respectively. As expected, a high-angle DNA bend correlated with impaired transcriptional activation. However, in this investigation, mutations designed to yield an active conformation failed to generate constitutive gene expression. Similarly, DNA bending alone is insufficient to explain the effects of operator mutations on the CysB-dependent transcription of *cysK* (Hryniewicz & Kredich, 1994, Monroe *et al.*, 1990). While DNA bending appears to be important in LysR-type regulation, the exact mechanism(s) by which bending and ligand-mediated conformational changes activate transcription are not yet fully understood.

Interactions with effectors

While most LTTRs interact with small-molecule effectors to activate transcription of their target genes, there are many interesting and atypical variations on this theme. OxyR, for example, is a redox-sensitive protein that activates expression of antioxidant defense genes in response to oxidative stress (Choi *et al*., 2001, Kullik *et al*., 1995). OxyR undergoes a conformational shift mediated by the formation of an intramolecular disulfide bond. Only the oxidized form is capable of activating transcription of its target genes, and differential gene regulation is associated with the remodeling of OxyR-DNA contacts.

There are other LTTRs with no apparent requirement for a direct interaction between an effector and the regulatory domain. However, unlike the case of OxyR where modification of the regulatory domain can substitute for the role of an effector, there are examples with no known alteration in this domain. Regulators in this category include the nitrogen assimilation control protein (NAC), the PhcA global regulator of virulence in a plant pathogen, and SyrM, which is involved in symbiotic nodule formation in leguminous plants (Clough *et al.*, 1997, Rosario & Bender, 2005, Swanson *et al.*, 1993). In these cases, the triggers that modulate expression of the target genes may include factors such as the conditions and extent of production of the LysR-type protein, the nature of interactions with specific DNA sequences, and/or the oligomeric state of the regulator. Variants of NAC that fail to repress one of its target genes also fail to form a tetramer, implicating this type of oligomer as a required feature of the repression mechanism (Rosario & Bender, 2005). Interestingly, these variants maintain the ability to bind DNA, to interact with RNA polymerase and to activate transcription at other loci. Moreover, these three activities appear to require only the N-terminal 86 residues of this LTTR.

Many LTTRs repress transcription at some loci and activate expression at others. The binding of a small-molecule effector is often needed for repression as well as for activation. For example, *N*-acetylserine, an effector that enables CysB to activate transcription at some loci, enhances CysB-mediated repression of *hsIJ* (Jovanovic *et al.*, 2003). Recently, heme was identified as an effector that enables one LTTR, HbrL, to repress *hem* genes needed for tetrapyrrole biosynthesis (Smart & Bauer, 2006). HbrL is also able to activate some *hem* loci, although this activation occurs in the absence of heme.

LTTR-effector interactions can influence different steps in transcription initiation. Ligand binding can alter DNA-binding affinity, promoter architecture, oligomerization and interactions

with RNA polymerase. In studies of flavonoid-induced gene regulation by NodD1, effector binding increased the affinity of the LTTR for *nod* gene promoters. However, this increase could be mediated by non-inducing (competing) flavonoids as well as by the specific flavonoid required for transcriptional activation (Peck *et al*., 2006). Therefore, transcriptional activation in this system requires the effector to regulate two separate steps: DNA binding and transcription initiation. The presence of multiple regulatory steps may be an inherently important aspect of effector-dependent transcription (Peck *et al*., 2006). The effector is likely to shift the equilibrium between LTTR conformations within a protein population toward a form it stabilizes. In a single equilibrium between an inactive and active form, the active form will occur at low frequency in the absence of effector. In a multi-step process, the chances diminish of attaining the transcriptionally-competent conformation without effector, and ligand binding should be required to shift the protein through more than one conformational change (Buck & Rosen, 2001, Ma *et al.*, 2002). While this model is consistent with regulatory studies, the exact nature of the protein conformation required for LTTR-activated transcription remains unclear.

Interactions with RNA polymerase

Several studies have addressed the specific interactions that occur between LTTRs and RNA polymerase (RNAP) subunits. In general, bacterial transcriptional activators directly contact the α and/or σ^{70} subunits of RNAP, as previously reviewed (Dove & Hochschild, 2005). The types of subunit contact tend to correlate with the positions of the regulator-DNA binding sites. For example, activators that bind positions overlapping or abutting the –35 region of the promoter typically contact an area of the σ^{70} subunit, designated region 4 (class II promoters). Activators that bind upstream of the RNAP binding site typically contact the C-terminal domain

(CTD) of the α subunit (class I promoters). For LTTRs, the exact position of DNA binding sites tends to vary not only for different regulators, but also for the same protein at different promoters. Therefore, individual studies are needed to clarify interactions between LTTRs and RNAP subunits and also to identify the surface(s) of these activators that make direct contact with RNAP, known as the positive control surfaces or the activating regions.

Protein-protein interactions between several LTTRs and α-CTD have been indicated by experimental approaches using a wide array of mutant alleles of *rpoA*, which encodes the α subunit of RNAP (Fritsch *et al.*, 2000, Lochowska *et al*., 2004, McFall *et al.*, 1998, Stauffer & Stauffer, 2005, Tao *et al.*, 1995). Studies of MetR at two different promoters, *metE* and *metH*, localize residues on two distinct faces of the α-CTD that are essential for MetR-activated transcription from one or both promoters (Fritsch *et al*., 2000). One of these distinct faces is only critical for *metE* activation. Moreover, within the cluster of residues of the second face of the α subunit, some residues affect activation of both genes, while others only affect the activation of either *metE* or *metH*. The number and location of MetR binding sites differ for the two promoters, suggesting that specific interactions depend on the position of the DNA-bound LTTR relative to the RNAP binding site. This conclusion is further supported by orientation studies demonstrating that either of the two α-CTD regions of RNAP is sufficient for *metE* activation, whereas *metH* activation involves the preferential interaction of MetR with one of the two asymmetrically positioned α-CTD regions (Fritsch *et al*., 2000). This preference presumably derives from rotational constraints that restrict the flexible α-CTD from wrapping around the DNA when MetR is on a helical face different from that of the remaining RNAP, as is the case for *metH* but not *metE*.

In recent studies of another LTTR, GcvA-activated transcription was shown to involve interactions with the α or σ^{70} subunit of RNAP depending on the promoter (Stauffer & Stauffer, 2005). In the region controlling expression from the *gcvB* promoter, the binding of GcvA to a single region from –29 to –76 relative to the transcriptional start site is characteristic of a class II promoter. Consistent with this class of promoter, interactions between the activator and σ^{70} at the *gcvB* promoter were indicated by the effects of RNAP variants that individually contain different alanine substitutions in the σ^{70}-CTD. While 5 of the 16 variants tested reduced *gcvB* expression, none significantly affected GcvA-dependent transcription of the *gcvTHP* operon. The architecture of the *gcvT* promoter is complex and involves multiple regulatory proteins. The closest GcvA binding site required for transcription from this promoter is positioned from –241 to –214 relative to the transcriptional initiation site. Consistent with the class III structure of this promoter, expression studies with two *rpoA* alleles encoding CTD deletions indicate that GcvA interacts with the α-CTD region to activate expression of *gcvT*.

It is likely that different interactions with RNAP subunits also distinguish the CysB-mediated control of several promoters (Lochowska *et al*., 2004). Whereas *cysP* is expressed from a class I promoter, the CysB binding sites used to activate the transcription of *cysJ* and *cysK* are characteristic of class II promoters. Consistent with these different promoter architectures, a mutant *rpoA* allele appears to reduce the CysB-activated expression of *cysP* but not *cysJ* or *K*. While detailed analyses of protein-protein interactions at the *cysJ* and *cysK* promoters remain to be reported, the CysB interactions with α-CTD that control *cysP* transcription have been extensively explored (Lochowska *et al*., 2004).

An activating region of CysB was identified that includes three residues in the turn region of the HTH DNA-binding motif, residues 27-30. These surface exposed residues are not critical

for DNA binding or for effector response, yet they exert positive control of transcription (Lochowska *et al*., 2004). Residue 27 corresponds to the position of a positive control substitution that similarly affects the ability of GcvA to activate both the *gcvB* and *gcvT* promoters (Jourdan & Stauffer, 1998, Stauffer & Stauffer, 2005). The possibility is raised that this turn region of the highly conserved DNA-binding motif may serve a common positive control function in LTTRs. Footprinting experiments with CysB and RNAP at the *cysP* promoter suggest that the activating region may be involved in positioning RNAP on the DNA (Lochowska *et al*., 2004). Studies of OxyR also demonstrate that physical interactions between LTTRs and RNAP are important for correctly positioning a protein complex to initiate transcription. RNAP is able to affect the footprint and transcriptional activity of a variant protein, OxyR H198R, which alone yields a reduced-form footprint under oxidizing conditions. RNAP interactions stabilize an active conformation of this variant that allows gene expression and produces a footprint characteristic of the active form of this LTTR (Kullik *et al.*, 1995). Moreover, this result indicates the reciprocity of protein-protein interactions, since OxyR increases RNAP binding in a cooperative fashion (Tao *et al*., 1995, Choi *et al*., 2001).

To determine the specific interactions between CysB and the α subunit of RNAP at the *cysP* promoter, alanine substitutions at residues 255-329 of α-CTD were tested for their effects on gene expression (Lochowska *et al*., 2004). A regulatory model based on these results suggests that the DNA-binding determinant of α-CTD, known as the 265 determinant (Gourse *et al.*, 2000), interacts with DNA in a non-sequence dependent fashion adjacent to the CysB-binding site. CysB-RNAP contacts involve a different region of α-CTD, known as the 273 determinant, which was previously shown to interact with the Fis regulatory protein (Bokal *et al.*, 1997). Residues in this region of α-CTD important for CysB regulation include K271, E273, H276, and

D280. To test this model, a LexA-based two hybrid system was used to investigate protein-protein contacts (Dmitrova *et al.*, 1998). Results from this approach support the conclusion that the wild-type CysB interacts with α-CTD and that specific interactions occur between the activating region of CysB (residues 27-30) and the 273 determinant of α-CTD (Lochowska *et al*., 2004). Further studies are needed to determine whether these results can be generalized for additional LTTR-RNAP interactions at class I promoters.

Known and Predicted LTTRs in *A. baylyi* ADP1

By considering the general features and sequences of LTTRs, we sought to assess the functions of 44 open reading frames (ORFs) likely to encode LysR-type regulators in strain ADP1 (Table 2.1). The putative LTTRs are listed by their unique numeric identifiers, with the prefix ACIAD, according to the genome sequence annotation (Barbe *et al*., 2004). These ACIAD identifiers are assigned consecutively to coding sequences in a clockwise direction from the origin of replication. The numeric designations shown in bold, 1 to 44, correspond to the positions in a list of putative LTTRs derived from ExtraTrain analysis, based on the criteria for the InterPro LysR family, IPR000847 (Mulder *et al.*, 2005, Pareja *et al*., 2006). As indicated in the table, 45% of these genes are oriented in a clockwise direction [+], and 66% are divergently transcribed from an adjacent gene.

The sequence similarity of the deduced amino acid sequence of each predicted LTTR was determined by comparison to database sequences using the basic local alignment search tool (BLAST) from the NCBI server (Altschul *et al.*, 1990). To limit the comparisons to proteins of known function, the Swiss-Prot protein database was used as the search target because these entries are carefully analyzed and annotated with literature-based curation (Apweiler *et al.*,

2004). With the exception of ACIAD1745 and ACIAD1979 (Table 2.1, nos. 9 and 11), significant sequence similarity was detected with other LTTRs in alignments that extended through both the DNA-binding domain and the regulatory domain. In most cases the similarity regulator. In several cases, the E-value in comparison with a known LTTR was less than 1^{-50}, a value sufficiently small to raise the suggestion of common function. As described next, genes in this category were examined further.

What's in a name?

ACIAD2597 (Table 2.1, no. 14) appears to be transcribed at the end of a cluster of genes involved in sulfur assimilation. This sequence has been annotated as Cbl, based on the high level of similarity with a regulatory protein involved in a complex regulon for sulfur metabolism and cysteine biosynthesis in *E. coli* and some other bacteria (Iwanicka-Nowicka & Hryniewicz, 1995). The structure of the regulatory domain of Cbl was recently characterized (Stec *et al*., 2006). However, several factors complicate a direct comparison of function. In *E. coli*, Cbl function is interconnected with that of CysB, a paralog that serves as a master regulator whose many roles include controlling *cbl* expression (van der Ploeg *et al.*, 2001, Bykowski *et al.*, 2002). Cbl and CysB share a high level of sequence identity that is not matched by ACIAD2597 and any other ADP1 open reading frame. Therefore, it is unlikely that the same dual regulatory Cbl-CysB system exists in ADP1.

Table 2.1 Possible LTTRs encoded by the *Acinetobacter baylyi* genome

Putative LTTRs in ADP1				**Possible Regulatory Target(s) of ADP1 LTTR**		
ACIAD[a] [direction][b] (size)	Similar LTTRs indentified by BLAST[c]	Sequence Identity to known LTTR[d]	E value[e]	Adjacent ORFs ACIAD	Possible function[f]	Possible LTTR binding sites, location[g]
0202 [+] (315 aa), **1**	P10183|NAHR_PSEPU P50324|NODD2_BRAEL	88/297, 29% 90/299, 30%	1 e-32 3 e-32	0196 to 0207 [+]	Twelve ORFS, varied functions	ATGC-N7-GCAT in 3' region of LTTR CDS
0434 [+] (299 aa), **2**	P67662|AAER_ECOLI P72131|PTXR_PSEAE	71/288, 24% 78/292, 26%	3 e-26 7 e-24	0433 [-] divergent	Short-chain dehydrogenase	TTTA-N7-TAAA in LTTR CDS
0813 [+] (307 aa), **3**	Q47141|HCAR_ECOLI P20668|GLTC_BACSU	85/274, 31% 80/249, 32%	2 e-28 5 e-28	0812 [-] divergent	Transport of inorganic ion (chromate, ChrA)	ATAC-N7-GTAT, ATTC-N7-GAAT; GTTG-N7-CAAC in LTTR CDS
0967 [+] (297 aa), **4** LysG/ArgP	Q9HW38|ICIA_PSEAE P94632|LYSG_CORGL	104/293, 35% 93/291, 31%	3 e-47 8 e-37	0966 [-] divergent	Efflux permease, amino acid export	ATCA-N7-TGAT, between divergent CDS; TTTA-N7-TAAA in LTTR CDS
1230 [+] (309 aa), **5**	P67662|AAER_ECOLI P72131|PTXR_PSEAE	95/313, 30% 87/297, 29%	4 e-29 7 e-27	1227 [+] 1229 [+]	Phosphatase, SurE Lipoprotein, NlpD	ATAG-N7-CTAT, CTTT-N7-AAAG, upstream of *nlpD*
1300 [+] (338 aa), **6**	P42722|CFXR_RALEU P45105|CYSB_HAEIN	67/259, 25% 63/256, 24%	5 e-11 6 e-11	1299 [+] 1301 [+]	Agmatinase, SpeB Glycerate dehydrogenase	TTAT-N7-ATAA, upstream of *speB*; ATGA-N7-TCAT in LTTR CDS
1425 [+] (296 aa), **7** SalR	P10183|NAHR_PSEPU P52679|PCPR_SPHCR	95/288, 32% 83/295, 28%	6 e-37 7 e-29	1424 [+]	SalA, salicylate 1-monooxygenase	TTAA-N7-TTAA, CTCA-N7-TGAG, upstream of *salA*
1675 [+] (299 aa), **8**	P20668|GLTC_BACSU P27111|CYNR_ECOLI	69/248, 27% 81/287, 28%	2 e-22 2 e-21	1674 [-] divergent	Transporter	TTTA-N7-TAAA, ATTC-N7-GAAT; ATAT-N7-ATAT in LTTR CDS
1745 [+] (299 aa), **9**	No strong similarity in the regulatory domain			1744 [-] divergent	Aspartate ammonia lyase, AspA	ATGC-N7-GCAT, between divergent CDS
1887 [+] (309 aa), **10**	P10183|NAHR_PSEPU Q02876|NODD1_RHITR	74/295, 25% 76/277, 27%	4 e-13 5 e-12	1886 [-] divergent	Acyl-CoA dehydrogenase	ATGA-N7-TCAT in LTTR CDS

Putative LTTRs in ADP1				**Possible Regulatory Target(s) of ADP1 LTTR**		
ACIAD[a] [direction][b] (size)	Similar LTTRs indentified by BLAST[c]	Sequence Identity to known LTTR[d]	E value[e]	Adjacent ORFs ACIAD	Possible function[f]	Possible LTTR binding sites, location[g]
1979 [+] (283 aa), **11**	No strong similarity in regulatory domain			1978 [-] divergent	Hypothetical protein	CTGT-N7-ACAG in LTTR CDS
2511 [+] (306 aa), **12**	P20668|GLTC_BACSU Q06610|RBCR_THIFE	82/292, 28% 65/239, 27%	4 e-20 2 e-17	2510 [-] divergent 2513 [+]	Symport/transporter Acetoacetyl-CoA transferase	TTTT-N7-AAAA, ATAA-N7-TTAT, TTAA-N7-TTAA; ATCA-N7-TGAT, overlaps LTTR CDS
2533 [+] (288 aa), **13**	P52685|MAUR_PARDE P25543|IRGB_VIBCH	77/285, 27% 72/243, 29%	3 e-20 4 e-15	2531 [-] divergent 2534 [+]	Arabinose transporter Ammonium transporter	TTTC-N7-GAAA, TTAT-N7-ATAA; CTAA-N7-TTAG, ATTC-N7-GAAT upstream of 2534
2597 [+] (307 aa), **14** <u>Cbl/CysB</u>	Q08598|CBL_KLEAE Q47083|CBL_ECOLI	162/309, 52% 154/302, 50%	1 e-87 7 e-90	2591-2597 [+]	Proteins for sulfur assimiliation: CysP, CysT, CysW, CysA, Cbl	ATAA-N7-TTAT, ATAT-N7-ATAT upstream of *cysP*; CTTT-N7-AAAG in *cyst*; GTGC-N7-GCAC in *cysA*; ATCC-N7-GGAT in *cbl*
2727 [+] (295 aa), **15**	P67662|AAER_ECOLI P52689|LTRA_KLEPN	95/290, 32% 87/295, 29%	9 e-35 1 e-34	2725 [-] divergent	NADH-dependent oxidoreductase	ATCC-N7-GGAT; TTTT-N7-AAAA in LTTR CDS
3129 [+] (309 aa), **16**	P67662|AAER_ECOLI P52689|LTRA_KLEPN	87/285, 30% 73/258, 28%	5 e-32 6 e-28	3128 [-] divergent 3130 [+]	Oxidoreductase Glutathione S-transferase	ATTG-N7-CAAT, GTTT-N7-AAAC, ATAT-N7-ATAT, ATTT-N7-AAAT
3225 [+] (299 aa), **17**	O68014|BENM_ACIAD P07774|CATM_ACIAD	88/299, 29% 90/302, 29%	4 e-21 6 e-20	3226 [+] 3222, 3223 [+]	Flavohemo-protein Cystathione beta-lyase, hypothetical protein	CTGC-N7-GCAG, GTCA-N7-TGAC, in 3222; ATCT-N7-AGAT upstream of LTTR
3274 [+] (284 aa), **18**	P94501|GLTR_BACSU Q44311|SOXR_ARTST	92/238, 3% 90/260, 34%	2 e-43 1 e-38	3273 [-] divergent 3276 [+]	Permease/transporter Hypothetical protein	ATAA-N7-TTAT, TTCA-N7-TGAA between divergent CDS; CTTG-N7-CAAG, upstream 3276

Putative LTTRs in ADP1				**Possible Regulatory Target(s) of ADP1 LTTR**		
ACIAD[a] [direction][b] (size)	Similar LTTRs indentified by BLAST[c]	Sequence Identity to known LTTR[d]	E value[e]	Adjacent ORFs ACIAD	Possible function[f]	Possible LTTR binding sites, location[g]
3629 [+] (301 aa), **19**	P71318|OXYR_PECCC Q04778|ALSR_BACSU	64/288, 22% 64/241, 26%	4 e-14 3 e-12	3628 [+] 3629 [+]	Uracil transporter Itself	CTTC-N7-GAAG, upstream 3628
3651 [+] (308 aa), **20**	P52689|LTRA_KLEPN P72131|PTXR_PSEAE	84/294, 28% 84/292, 28%	4 e-27 7 e-24	3650 [-] divergent	Hypothetical protein	TTGC-N7-GCAA, LTTR CDS
0105 [-] (299 aa), **21**	P10183|NAHR_PSEPU P52679|PCPR_SPHCR	91/295, 30% 88/295, 29%	8 e-43 3 e-34	0106 [+] divergent	L-lactate permease (IldP)	ATTG-N7-CAAT, between divergent CDS
0323 [-] (298 aa), **22**	P39127|CITR_BACSU P44418|OXYR_HAEIN	61/269, 22% 68/284, 23%	9 e-12 7 e-09	0324 [-] 0323 [-]	Hypothetical protein Itself	TTTT-N7-AAAA, GTTT-N7-CAAA, upstream; TTCA-N7-TGAA, in LTTR CDS
0448 [-] (298 aa), **23**	Q47141|HCAR_ECOLI Q04778|ALSR_BACSU	120/286, 41% 104/286, 36%	4 e-51 4 e-44	0449 [+] divergent 0446 [-]	Cation transporter Hypothetical protein	TTTA-N7-TAAA, in LTTR CDS; ATAT-N7-CTAT, ATAA-N7-TTAT, between divergent CDS
0457 [-] (309 aa**), 24**	P27111|CYNR_ECOLI P20668|GLTC_BACSU	60/239, 25% 72/276, 26%	7 e-20 4 e-17	0459 [+] divergent	Demethylase-oxygenase	ATGT-N7-ACAT, in LTTR CDS; TTCT-N7-AGAA, between divergent CDS
0461 [-] (293 aa), **25**	P0A9F3|CYSB_ECOLI P45600|CYSB_KLEAE	76/274, 27% 61/227, 26%	9 e-20 2 e-19	0460 [-] 0462 [-]	Hypothetical protein Hypothetical protein	TTAA-N7-TTAA, in LTTR CDS; TTTT-N7-AAAA, ATAA-N7-TTAT upstream 0462
0746 [-] (295 aa), **26**	Q06610|RBCR_THIFE P25544|RBCR_CHRVI	92/288, 31% 94/292, 32%	8 e-30 2 e-29	0747 [+] divergent	Ferridoxin NADP+ reductase	CTTT-N7-AAAG in LTTR CDS

Putative LTTRs in ADP1				**Possible Regulatory Target(s) of ADP1 LTTR**		
ACIAD[a] [direction][b] (size)	Similar LTTRs indentified by BLAST[c]	Sequence Identity to known LTTR[d]	E value[e]	Adjacent ORFs ACIAD	Possible function[f]	Possible LTTR binding sites, location[g]
1063 [-] (301 aa), **27** OxyR, EstR	P0ACQ4|OXYR_ECOLI Q9X725|OXYR_ERWC	99/290, 34% 97/290, 33%	9 e-31 2 e-30	1066 to 1063 [-]	Rubredoxin (RubA), reductase (RubB), esterase (EstB] and LTTR OxyR	Operon is expressed constitutively (Geissdorfer *et al*., 1999)
1082 [-] (329 aa), **28**	P10183|NAHR_PSEPU P55619|SYRM1_RHISN	91/288, 31% 89/285, 31%	3 e-35 2 e-33	1084 [+] divergent	Isocitrate lyase (AceA); unknown ORF may exist between LTTR and AceA	ATTA-N7-TAAT, ATTT-N7-AAAT; ATTT-N7-AAAT, TTCT-N7-AGAA in LTTR
1152 [-] (286 aa), **29**	Q44311|SOXR_ARTST P94501|GLTR_BACSU	77/290, 26% 77/287, 26%	3 e-18 6 e-18	1154 [+] divergent 1151 [-]	Transport protein Argininosuccinate synthase (ArgG)	ATCA-N7_TGAT, between divergent CDS; TTCA-N7-TGAA in LTTR; TTAG-N7-CTAA upstream *argG*
1294 [-] (295 aa), **30**	gb|AAG04811.1|GbuR P42722|CFXR_RALEU	179/284, 63% 52/195, 26%	7 e-10 4 e-10	1295 [+] divergent	Hypothetical protein	ATCT-N7-AGAT, ATTT-N7-AAAT, ATAA-N7-TTAT
1435 [-] (304 aa), **31** BenM	P07774|CATM_ACIAD O33945|CATR_ACILW	178/294, 60% 131/306, 60%	8 e-100 4 e-61	1436 [+] *benA* divergent	BenM-CatM regulon discussed in text	ATAC-N7-GTAT, known binding site
1445 [-] (303 aa), **32** CatM	O68014|BENM_ACIAD O33945|CATR_ACILW	178/294, 60% 118/297, 39%	8 e-100 2 e-55	1446 [+] *catB* divergent	BenM-CatM regulon discussed in text	ATAC-N7-GTAT, known binding site
1539 [-] (309 aa), **33**	Q47005|NAC_ECOLI Q08597|NAC_KLEAE	93/312, 29% 89/309, 28%	2 e-25 6 e-22	1540, 1539, 1537 [-]	Substrate-binding protein, LTTR similar to ACIAD1543, Hypothetical	CTTG-N7-CAAG, upstream ACIAD1540; AATT-N7-TTAA, TTGT-N7-ACAA, CTTT-N7-AAAG, in LTTR
1543 [-] (309 aa), **34**	Q47005|NAC_ECOLI Q08597|NAC_KLEAE	76/232, 32% 74/252, 29%	4 e-23 6 e-21	1545, 1543, 1541 [-]	Hypothetical protein, LTTR similar to ACIAD1539	TTGA-N7-TCAA, GTCT-N7-AAAC, ATTC-N7-GAAT, upstream ACIAD1545; TTTC-N7-AAAC upstream of *citA*

Putative LTTRs in ADP1				**Possible Regulatory Target(s) of ADP1 LTTR**		
ACIAD[a] [direction][b] (size)	Similar LTTRs indentified by BLAST[c]	Sequence Identity to known LTTR[d]	E value[e]	Adjacent ORFs ACIAD	Possible function[f]	Possible LTTR binding sites, location[g]
1602 [-] (294 aa), **35**	P52685|MAUR_PARDE P45099|GCVA_HAEIN	70/278, 25% 73/291, 25%	1 e-15 2 e-09	1604 [+] divergent	Oxidoreductase (MmsA), most likely first in an operon	No perfect match in LTTR CDS or between divergent CDS
1656 [-] (305 aa), **36**	P52689|LTRA_KLEPN P72131|PTXR_PSEAE	87/293, 29% 86/293, 29%	3 e-32 4 e-29	1656 [-] 1657 [-]	Itself Hypothetical protein	ATTA-N7-TAAT, upstream LTTR CDS; GTTT-N7-AAAC, upstream ACIAD1657
1659 [-] (315 aa), **37**	P0A9F6|GCVA_ECOLI P45099|GCVA_HAEIN	99/318, 31% 98/317, 30%	9 e-35 2 e-28	1660 [+] divergent	Glutaryl-CoA dehydrogenase (GcdH)	ATTA-N7-TAAT, ATCT-N7-AGAT, between divergent CDS
1762 [-] (306 aa), **38** MdcR	P52684|MAUR_KLEPN P27111|CYNR_ECOLI	128/301, 42% 65/257, 25%	9 e-67 3 e-19	1753 [+] 1762 [+]	1753-1761, Mdc operon Itself	ATAA-N7-TTAT, TTAA-N7-TTAA, upstream of *mdcA*; ATTG-N7-CAAT in LTTR CDS
2384 [-] (306 aa), **39** MetR	P0A9G1|METR_SHIFL P0A2Q5|METR_SALTI	122/291, 41% 123/291, 42%	4 e-60 1 e-59	2385 [+] divergent	Permease MFS superfamily	CTCA-N7-TGAG, ATGA-N7-TCAT, between divergent CDS
2522 [-] (309 aa). **40**	P94501|GLTR_BACSU P0A2Q5|METR_SALTI	71/285, 24% 79/295, 26%	2 e-17 2 e-16	2523 [+] divergent 2521 [-]	GntR family regulator Hypothetical protein	TTTT-N7-AAAA, between divergent CDS; ATAA-N7-TTAT, ATAT-N7-ATAT upstream ACIAD2521
2742 [-] (318 aa), **41**	P52689|LTRA_KLEPN P72131|PTXR_PSEAE	159/293, 54% 89/298, 29%	4 e-85 1 e-30	2743 [+] divergent	Amino acid efflux transmembrane protein	TTTT-N7-AAAA, TTTC-N7-GAAA, in LTTR CDS
3169 [-] (303 aa), **42**	Q04778|ALSR_BACSU P52666|BUDR_KLETE	100/287, 34% 101/286, 35%	3 e-39 4 e-36	3170 [+] divergent	Transporter	ATAT-N7-ATAT overlaps start of LTTR CDS
3490 [-] (328 aa), **43**	P20668|GLTC_BACSU P27111|CYNR_ECOLI	73/298, 24% 79/290, 27%	5 e-13 3 e-12	3493, 3492 [-]	Hypothetical protein	TTTA-N7-TAAA in LTTR CDS
3615 [-] (302 aa), **44**	P72131|PTXR_PSEAE P52689|LTRA_KLEPN	86/293, 29% 82/288, 28%	2 e-29 3 e-26	3616 [+] divergent	Aldehyde reductase (AlrA)	No perfect match in LTTR or between divergent CDS

Notes to Table 2.1

[a]ACIAD identifiers correspond to unique coding sequences (CDS) in the ADP1 genome (Barbe *et al.*, 2004). Designations of known or possible (underlined) regulators are discussed in the text.

[b]A clockwise [+] or counterclockwise [-] direction indicates the chromosomal orientation of each CDS (Barbe *et al.*, 2004).

[c]The basic local alignment search tool (BLAST) was used from the NCBI server (Altschul *et al.*, 1990). Similar proteins are indicated by their primary accession number and entry name from the Swiss-Prot entry.

[d]Expressed as the number of identical residues in a pairwise alignment divided by the total number of residues in the local alignment.

[e]The expectation (E) value provides an indication of the likelihood of achieving a similar alignment score merely by chance. The greater the significance of the match, the smaller the E-value.

[f]Assigned by the genome sequence annotation (Barbe *et al.*, 2004).

[g]Identified with the Pattern Locator program (Mrazek and Xie, 2006).

The genetic organization of all *cys* genes is strikingly different in ADP1 and *E. coli*. The putative *cbl* gene of ADP1 does not have the typical divergent gene location from a target. As indicated in Table 2.1, it is the last of a gene cluster, ACIAD2591-2597, which appears to form an operon involved in sulfate-thiosulfate transport. This position suggests that it might control expression from the *cysP* promoter. In *E. coli*, CysB rather than Cbl regulates an operon that initiates with *cysP*. The genetic composition of this *E. coli* operon is similar but not identical to that in ADP1. Notably, the *cysM* gene involved in cysteine biosynthesis, is part of the *cysP* operon in *E. coli*. In ADP1, the gene encoding CysM (ACIAD3071), cysteine synthase, is distal from the putative *cysP* operon. Whether ACIAD2597 should be designated Cbl, CysB, or something else, a role in regulating sulfur assimilation is suggested. However, the regulation of cysteine biosynthesis and the control circuit for sulfur metabolism in ADP1 are likely to differ significantly from that of *E. coli*.

Similar difficulties arise with assigning a possible function to ACIAD2384 (Table 2.1, no. 39). Sequence similarity leads to the tentative designation of MetR. In *E. coli* and *Salmonella enterica* serovar Typhimurium, MetR is a homocysteine-responsive LTTR that regulates methionine biosynthesis by controlling expression of the *metE* and *metH* genes (Fritsch *et al*., 2000). In these bacteria, *metR* is divergently transcribed from *metE*, which is separated on the chromosome from *metH* by more than 200 kbp. In ADP1, the potential *metR* is not located near the genes encoding MetE (ACIAD3523) or MetH (ACIAD1045). The ACIAD2384 gene is divergently transcribed from one encoding a putative permease of the multifacilitator superfamily (ACIAD2385). The divergent location might suggest a regulatory connection between ACIAD2384 and ACIAD2385. However, the size of the intergenic region, 533 nt, is larger than those between most LTTRs and their divergent targets, which tend to be in the range

of 100-300 nt. Only experimental studies will be able to establish whether ACIAD2384 plays a role in methionine biosynthesis.

ACIAD1294 (Table 2.1, no. 30) corresponds to another putative LTTR with significant sequence similarity to a known regulator, GbuR (locus tag PA1422) of *Pseudomonas aeruginosa* (Nakada & Itoh, 2002). The target of this regulator is the divergently transcribed *gbuA*, which encodes an enzyme in the arginase/agmanitase family involved in arginine catabolism. In contrast, the divergently transcribed gene from the ADP1 homolog encodes a hypothetical gene of unknown function. Nevertheless, in the genomic vicinity of the ACIAD1294 LTTR gene is a *gbuA*-like gene, corresponding to ACIAD1299. The putative ADP1 GbuA (sometimes designated SpeB) is 88% identical in sequence to that of *P. aeruginosa*. The nearest regulatory gene to the ADP1 *gbuA*-like gene is immediately downstream, and it encodes another putative LTTR, ACIAD1300 (Table 2.1, no. 6). The sequence of this latter regulator does not closely resemble that of GbuR, and it is not clear if either ACIAD1294 or ACIAD1300 regulates *gbuA* expression. Moreover, the metabolic function of *gbuA* (ADIAD1299) remains to be established in ADP1.

The regulatory and metabolic functions of the Mdc proteins, including MdcR (ACIAD1762; Table 2.1, no. 38), can be inferred with greater confidence because of synteny and previous biochemical studies. MdcR, also known as MauR, regulates malonate catabolism in *Klebsiella pneumoniae* (Peng *et al.*, 1999). The *mdcR* gene of *K. pneumoniae* is located convergently at the 3' end of an *mdc* operon, as it is in ADP1. While this is not the typical LTTR genetic arrangement, MdcR functions as an activator of the *mdc* operon and as a negative autoregulator (Peng *et al*., 1999). The *mdcABCDEGHLM* operon of ADP1 is highly conserved with the corresponding one in *Pseudomonas putida* and other sequenced pseudomonad genomes.

The encoded proteins have been investigated in *P. putida* to confirm the activity of malonate decarboxylase and to indicate that MdcLM is a malonate transporter (Chohnan *et al.*, 1999).

Significant similarity between two additional predicted LTTRs in ADP1 and database entries provided few clues about their function. ACIAD2742 (Table 2.1, no. 41) resembles LtrA, the product of a gene identified only as a putative regulatory locus downstream of genes involved in citrate fermentation in *Klebsiella pneumoniae* (Bott *et al.*, 1995). ACIAD0448 (Table 2.1, no. 23) closely resembles HcaR of *E. coli*, a regulator of the initial steps in the degradation of phenylpropanoid compounds such as 3-phenylpropionic acid (Diaz *et al.*, 1998). The ADP1 homolog is divergently transcribed from a putative cation transporter, and there are no genes in the vicinity with similar functions to those of the *E. coli* catabolic pathway. Further complicating nomenclature issues, HcaR (ACIAD1728) has previously been used to designate a novel member of the MarR family of regulators involved in the degradation of hydroxycinnamates by ADP1 (Parke & Ornston, 2003).

Problems with nomenclature also affect ACIAD0967 (Table 2.1, no. 4). Current ADP1 genome annotation indicates that this locus corresponds to IciA, a regulator of replication initiation in *E. coli* (Lee *et al.*, 1996). The *iciA* locus of *E. coli* has also been designated *argP*, based on its involvement in arginine export. Specifically, the ArgP (IciA) LTTR regulates expression of *argO* (also designated *yggA*), which encodes an arginine exporter similar to LysE of *Corynebacterium glutamicum* (Nandineni & Gowrishankar, 2004, Eggeling & Sahm, 2003). While the *argP* and *argO* genes are homologs of *lysG* and *lysE* of *C. glutamicum*, there appear to be differences in the functional specificity and regulation of the two systems (Nandineni & Gowrishankar, 2004). In ADP1, the divergent location of the genes for ACIAD0967 and ACIAD0966 matches that of *lysG* and *lysE*, respectively. Yet, the sequences of these ADP1

proteins more closely resemble ArgP and ArgO than LysG and LysE. A role in regulating amino acid export is suggested for the ACIAD0967 LTTR, although the specificity of the transport system cannot be inferred.

ACIAD1659 (Table 2.1, no. 37) has been predicted to be GcvA in some databases. This regulator in *E. coli* controls a glycine cleavage system, encoded by *gcvTHP* (Stauffer & Stauffer, 2005). However, this cleavage system does not appear to be encoded by the ADP1 genome. GcvA is an interesting LTTR in that its regulatory response to glycine does not appear to occur directly by binding this ligand. Rather, regulation is mediated through interactions with another regulator, the GcvR repressor (Heil *et al.*, 2002). There is no homolog for GcvR encoded by ADP1. Thus, while ACIAD1659 displays approximately 30% sequence identity to the GcvA protein of *E. coli*, it is unlikely to have the same function in strain ADP1.

Genetic context

Approximately one third of the genes proposed to encode LTTRs in ADP1 are not transcribed divergently from a possible target gene. Nevertheless, alternative genetic arrangements can also place LTTR-encoding genes in the vicinity of their targets. For example, as noted above, the convergently transcribed *mdc* operon is the likely target of MdcR (ACIAD1762; Table 2.1, no. 38) (Peng *et al*., 1999). In a different arrangement, SalR (ACIAD1425; Table 2.1, no. 7) regulates expression of *salA*, the gene immediately upstream of, and in the same direction as, its own (Jones *et al.*, 2000). In an attempt to identify possible regulatory targets, genes in the immediate vicinity of the predicted LTTR genes are indicated in Table 2.1. Many of these potential targets encode membrane proteins and transporters, consistent

with the importance of importing growth substrates for the saprophytic lifestyle of strain ADP1, as previously discussed (Barbe *et al*., 2004, Young *et al.*, 2005).

The prediction of regulatory targets should be interpreted cautiously. LTTRs also regulate distantly located genes, and some are known not to regulate any of their neighboring loci. Moreover, errors in genome annotation can be misleading. For example, the gene encoding ACIAD1063 (Table 2.1, no. 27) appears to be co-transcribed with its upstream gene, *estB*. Because of this genetic context, it was annotated *estR*. Nevertheless, a previous study indicates that on the basis of sequence, function, and regulation, this gene is more appropriately designated *oxyR* (Geissdorfer *et al.*, 1999). The amino acid sequence of ACIAD1063 aligns with those of well-characterized OxyR proteins and was shown to contain an 18-amino acid OxyR signature with invariant conservation of the two cysteine residues involved in regulation via disulfide bond formation (Geissdorfer *et al*., 1999). Consistent with its designation, *oxyR* was shown to affect sensitivity to hydrogen peroxide. Furthermore, expression of *oxyR* and its neighbors was investigated experimentally. Four genes, *rubA, rubB, estB*, and *oxyR* were shown to comprise an operon that appears to be expressed constitutively (Geissdorfer *et al*., 1999).

Genome position and sequence analysis suggest that two of the putative LTTRs, ACIAD1539 and ACIAD1543 (Table 2.1, nos. 33 and 34), are paralogs that arose from a relatively recent gene duplication event. In addition to the close position of their genes on the chromosome, separated by less than 3 kbp of DNA, their deduced amino acid sequences indicate a high level of conservation. In a full-length alignment of the two sequences, 57% of the residues are identical and 76% are similar. This similarity is reminiscent of that between the BenM and CatM paralogs. A more distantly related cousin of BenM and CatM, ACIAD3225 (Table 2.1, no. 17), was identified that shares approximately 29% identity with them. Multiple sequence analysis

indicates that BenM and CatM form a small subfamily of the putative ADP1 LTTRs that, in addition to ACIAD3225, includes ACIAD3169, ACIAD0813, and ACIAD0448 (Table 2.1, nos. 42, 3, and 23, respectively). Sequence similarity detected throughout the entire length of these proteins suggests common features of effector binding and response.

Binding sites

To detect potential binding sites for LTTRs in the ADP1 genome, a newly developed computer program was used, Pattern Locator (Mrazek & Xie, submitted). We searched for a characteristic LTTR-binding site comprised of a T-n11-A motif within an interrupted dyadic sequence (Schell, 1993). Table 2.1 shows exact matches to this pattern that occur in the vicinity of possible target genes or within LTTR-encoding genes where they could participate in negative autoregulation. Few LTTR-binding sites have been determined in ADP1, although one sequence matching the consensus, ATAC- n7-GTAT, is known to be important for the BenM-CatM regulon (Bundy *et al*., 2002, Clark *et al.*, 2002, Romero-Arroyo *et al*., 1995). While it is unclear whether the binding sequences in Table 2.1 are functional, these sites may serve as starting points for further experimental exploration. The exact locations of the sites, and patterns that match variations on this theme, can readily be determined with the Pattern Locator program (http://www.cmbl.uga.edu/software/patloc.html).

The BenM-CatM Regulon of *A. baylyi* ADP1

The most extensively studied LysR-type regulon in ADP1 involves the degradation of benzoate and related aromatic compounds via the β–ketoadipate pathway (Figure 2.1). The first step in understanding the regulation of benzoate degradation was the characterization of a mutant

selected by growth on β–ketoadipate as the sole carbon source (Neidle *et al*., 1989). Normally, this substrate is degraded only when the necessary enzymes, PcaIJF or CatIJF, are properly induced (Figure 2.1). In the wild-type strain, *catIJF* expression depends on CatM, a negatively-autoregulated LTTR that activates transcription from the *catA* and *catB* promoters in response to muconate (Romero-Arroyo *et al*., 1995). The regulatory mutant grows on β–ketoadipate because a CatM variant, CatM(R156H), activates transcription of the *catBCIJFD* operon in the absence of its effector, muconate. The variant similarly mediates muconate-independent *catA* expression.

In the first study of CatM, the redundant regulatory effects of BenM were not yet understood, and because of this, some results were misinterpreted (Neidle *et al*., 1989). It was later shown that BenM, as well as CatM, activates transcription from the *catA* and *catB* promoters in response to muconate. The discovery of this response to muconate was surprising. It was expected that BenM would respond to benzoate or a metabolite distinct from muconate. The divergent arrangement of *benM* and the *benABCDE* operon suggested that BenM would regulate the early steps in benzoate catabolism. The prediction that muconate would not be the primary effector for this regulator stemmed from the observation that the BenABC and BenD enzymes are not fully induced in muconate-grown cells (Neidle *et al.*, 1987).

Effector response by BenM and CatM

Transcriptional studies conducted *in vivo* and *in vitro* clarified the role of muconate and benzoate as effectors (Bundy *et al*., 2002, Collier *et al*., 1998, Ezezika *et al.*, 2006). Both compounds individually enable BenM to activate *benA* transcription. As a sole effector, muconate elicits higher levels of gene expression than does benzoate. However, the inability of muconate alone to induce maximal levels of the *ben* gene-encoded enzymes can be explained by

the combined effect of both compounds on transcription. In response to muconate and benzoate together, the BenM-dependent levels of *benA* expression are significantly higher than the sum of those due to each compound individually (Bundy *et al*., 2002, Collier *et al*., 1998). This type of transcriptional synergism by one regulator in response to two effectors is novel yet physiologically significant. Such regulation provides a rapid mechanism for integrating cellular signals in a complex pathway where the accumulation of intermediary metabolites can be toxic. Benzoate initially increases *ben*-gene expression, but maximal transcription occurs only when muconate signals that degradation is proceeding properly. As benzoate availability decreases, gene expression lowers, but muconate-regulated transcription allows substrate consumption to be completed.

The unusual response of BenM to muconate and benzoate raises interesting questions about ligand binding and response by this and other LTTRs. Comparisons between CatM and BenM provide a good model for addressing such questions. The greatest sequence similarity between these paralogs is in the region that corresponds to the wHTH motif of the DNA-binding domain. In this region, the 58 N-terminal residues of BenM and CatM are 84% identical and 95% similar. Therefore, it is not surprising that in the absence of effectors, BenM and CatM each bind the same sites of the *benA* promoter region (Sites 1 and 3, Figure 2.1) and yield nearly identical DNase I footprints (Bundy *et al*., 2002, Ezezika *et al*., 2006). However, CatM responds only to muconate and activates *benA* expression at a low level that is insufficient to support growth on benzoate as the sole carbon source in the absence of BenM. Similarly, CatM fails to respond to benzoate at its other target loci, such as the *catB* promoter.

In the EBD region of BenM and CatM, responsible for binding muconate by both regulators, the two sequences are 54% identical and 77% similar. This protein region, residue 81

to the C terminus, should be largely responsible for distinguishing the regulatory capabilities of BenM and CatM and for providing the structural basis of benzoate recognition by BenM. Effector-binding studies using fluorescence emission spectroscopy established that BenM has a higher affinity for muconate than for benzoate as the sole effector (Clark *et al.*, 2004b). However, comparable studies could not be done with CatM, because this technique detects interactions that affect the environment of tryptophan, an amino acid not found in CatM. The studies of full-length BenM and BenM-EBD showed that benzoate decreases the apparent affinity of the regulator for muconate (Clark *et al*., 2004b). This decrease is consistent with competitive binding of the two effectors, but the data do not reveal whether the effectors bind to the same or different sites on the protein. To improve our understanding of the protein-effector interactions of BenM and CatM, structural studies were conducted.

Protein fold of CatM-EBD and BenM-EBD

The structures of BenM-EBD and CatM-EBD, with and without their biological effectors benzoate and muconate, were determined by X-ray crystallography (Clark *et al.*, 2004a, Ezezika *et al*., 2007a). These proteins lack the first 80 N-terminal residues including the DNA-binding domain. The BenM-EBD structure consists of two α/β domains that use Rossmann folds separated by a hinge formed from the central regions of two antiparallel β-strands (Figure 2.2). Domain I includes residues 87 to 161 of the full-length protein, and a C-terminal helix, residues 268 to the end. Domain II contains the sequence between the regions of domain I, residues 162 to 267. The CatM-EBD structure is nearly identical to that of BenM-EBD. The positions of the alpha-carbon backbones throughout the entire two structures have a root mean square deviation of 1.4 Å. Two subunits associate to form a stable dimer in all the structures of BenM- and CatM-

EBD. The dimer is also the principle species of BenM-EBD in solution, even though the full-length protein is a tetramer in solution (Bundy *et al*., 2002, Clark *et al*., 2004a). The DNase I footprints indicate that the tetrameric species is biologically active (Bundy *et al*., 2002, Ezezika *et al*., 2006).

Effector-binding sites in BenM-EBD and CatM-EBD

Structural analyses of BenM-EBD crystals soaked with muconate demonstrate the binding of this effector in a pocket between domains I and II (Figure 2.2). This pocket, designated the primary effector-binding site, is similarly occupied by muconate in CatM-EBD or benzoate in BenM-EBD in structures from protein crystals soaked in muconate or benzoate, respectively (Ezezika *et al*., 2007a). The position of this site is consistent with a general regulatory model for LTTRs in which an effector binds in this interdomain pocket. The hinge-like β-strands allow the two regulatory domains of the protein to clamp down on the effector. Conformational changes that occur upon ligand binding can be propagated throughout the structure to affect interactions with promoter-region DNA and with RNAP to regulate transcription. While this model has been proposed previously (Muraoka *et al*., 2003, Smirnova *et al*., 2004, Tyrrell *et al*., 1997), structural studies of BenM- and CatM-EBD provide the first examples of inducer-bound LTTRs.

In BenM-EBD crystals soaked only in benzoate, molecules of this effector were identified in the primary effector-binding site and also in a secondary effector-binding site inside the hydrophobic core of domain I (Figure 2.2). Soaking CatM-EBD crystals in solutions with high concentrations of benzoate did not yield effector-bound structures of this regulator.

Consistent with its inability to respond to benzoate as an effector, CatM lacks key residues that interact with benzoate in the secondary binding site of BenM.

In the primary binding site, residues involved in muconate-regulator interactions are remarkably well conserved between BenM and CatM (Figure 2.3A). While most of the ligand-protein contacts are direct, water molecules bridge some interactions. The carboxylate atoms of the muconate form direct hydrogen bonds to side chain oxygen atoms of Ser99 and Thr128, and the main chain amide nitrogen atoms of Thr128, and Phe203 (Figure 2.3A). A hydrophobic pocket forms from the phenyl ring of Phe203 and the aliphatic side chains of Val97 and Arg146. Remarkably, CatM differs from BenM at only one amino acid in this binding site. Gly98 of BenM is replaced by a serine in this position of CatM.

Ser99, Thr128, Phe203, and the water-bridged Val227 are at the amino terminal ends of helices H1, αH3, αH6, and αH7, respectively (Figure 2.2). These four helices are oriented so that their positive dipole moments stabilize the negatively charged carboxyl groups of bound muconate. These interactions are significant, as there are no nearby positively charged amino acids. When Arg146, which is relatively distant from the muconate, forms a salt-bridge to Glu162, it is unable to stabilize the negative charge of muconate. In contrast to muconate, benzoate fits loosely in the primary binding site (not shown) (Ezezika *et al*., 2007a). The carboxyl group of benzoate interacts with the hydroxyl side chain of Ser99, and hydrophobic interactions come from Phe203, Leu147, and Arg146.

In the secondary-binding site of BenM-EBD (Figure 2.3B), benzoate forms a salt bridge with the side chain of Arg160. Hydrogen bonds form with the side chain hydroxyl of Tyr293 and the main chain amide nitrogen of Leu104 via the carboxyl group. The side chains of Leu100, Leu105, Ile108, Phe144, Leu159, Ile269, and Ile289 form a hydrophobic pocket appropriately

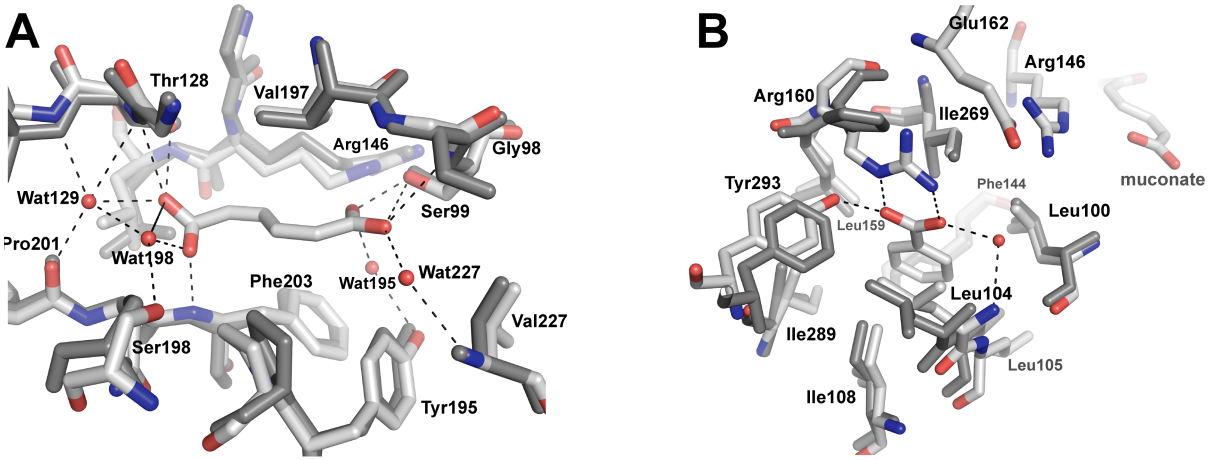


Figure 2.3: Views of the primary (A) and secondary (B) effector-binding sites of BenM-EBD and CatM-EBD. The structure of CatM-EBD (dark gray) was superimposed onto that of BenM-EBD (light gray) in the effector-binding regions using local residues for the docking. In (A), muconate is shown bound in the primary sites of both BenM and CatM. For clarity, only the water molecules coordinated to BenM are depicted. Water molecules (Wat) are numbered according to their nearest amino acid. Water molecules that coordinate to CatM are located in equivalent positions. The amino acids surrounding muconate are essentially the same in BenM and CatM except for Gly98, which corresponds to Ser98 in CatM. In (B), a benzoate molecule is in the secondary site of BenM only. CatM has water molecules in this region (not shown). While the hydrophobic residues lining the binding site for the phenyl ring are conserved, CatM appears unable to bind the carboxylate group.

sized for the phenyl ring. While the hydrophobic pocket is substantially conserved in CatM, residues of BenM that interact specifically with the carboxyl group of benzoate are absent in CatM. The side chain of Phe293 of CatM can not hydrogen bond to benzoate as Tyr293 does in BenM. Unlike Arg160 in BenM, His160 of CatM is too distant to interact directly with benzoate.

A model for transcriptional synergism

In one crystal structure of BenM-EBD, muconate and benzoate appear to be bound in the primary and secondary effector-binding sites, respectively (Figure 2.2) (Ezezika *et al*., 2007a). In the secondary site, benzoate most likely produces the synergistic response of BenM by accentuating the signal from muconate in the primary site. This process may involve two mechanisms, the bulk displacement of the local environment around the benzoate-binding site and/or a change in the electrostatic environment. In the secondary site, benzoate binds in the vicinity of H1, a helix that contacts muconate in the primary binding site (Figure 2.2). Glu162 appears to be a key residue that links regulatory effects from the two ligand-binding sites.

The interactions of Glu162 with its neighboring Arg160 or Arg146 can change the electrostatic distribution within BenM. When benzoate is bound in the secondary site, Arg160 forms a salt bridge with the carboxyl group of the benzoate (Figure 2.4). Under these conditions, Glu162 can form a salt bridge with Arg146. Neutralization of Arg146 enhances the negative charge of muconate and thus should draw the surrounding helices more strongly toward the muconate. This in turn may stabilize conformational changes needed for transcriptional activation. In the absence of benzoate, Glu162 can form a salt-bridge with Arg160. Under these conditions, the positively charged Arg146 in the vicinity of muconate might weaken the electrostatic pull between the effector and the surrounding helices. Such alterations could

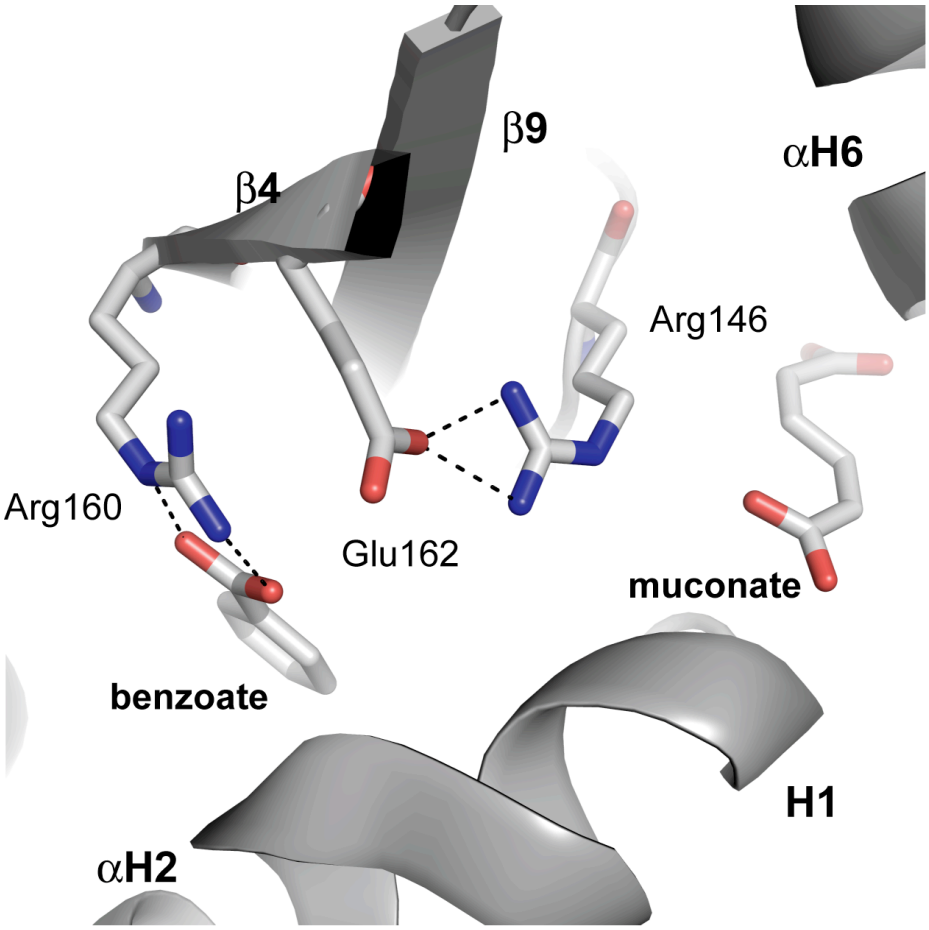


Figure 2.4: Residues in a charge relay system postulated to underlie the synergistic response of BenM to benzoate and muconate. Synergistic transcriptional activation may rely on the simultaneous occupation of the primary and secondary effector-binding sites by muconate and benzoate, respectively. Glu162 sits in the hinge-like region that forms the central portion of two β-strands (β4 and β9) connecting domains I and II. In the absence of benzoate, Glu162 is free to interact with either Arg160 or Arg146. In the presence of benzoate, the association between Arg160 and benzoate, permits Glu162 to interact only with Arg146. Therefore, the presence of benzoate diminishes the ability of the positively charged Arg146 to neutralize the negatively charged muconate. The net result of this shielding effect of Glu162 on Arg146 is to enhance the interactions of muconate with the helical dipole moments of H1, αH2, αH3, and αH6. This enhancement is predicted to increase conformational changes that draw the helices more closely toward muconate (Ezezika *et al*., 2007a; Craven *et al*., 2009).

account for conformational changes between a tetramer bound only to muconate and one bound to both effectors.

Glu162, which is conserved in BenM and CatM, is located in the middle of β-strand 4 in the hinge-like region between domains I and II. This region is critical for effector response by other LTTRs such as CysB (Lochowska *et al*., 2001), CbbR (Dangel *et al.*, 2005) and the distantly related, OccR (Akakura & Winans, 2002a). In some OccR variants, two distant amino acid substitutions activate inducer-independent transcription at higher levels than either substitution alone (Akakura & Winans, 2002a). In these OccR variants, the amino acid substitutions could be replicating the types of changes in charge distribution involved in the synergistic response of BenM.

Transcriptional regulation and conformational changes in BenM- and CatM-EBD

The binding of muconate to the primary effector-binding site of BenM-EBD causes small but significant conformational changes within a single subunit. Models of full-length BenM structure, based on the structure of CbnR, have been proposed and depict propagation of these seemingly minor changes through the tetramer to the DNA-binding domains (Ezezika *et al*., 2007a). The four DNA-binding domains of CbnR are located on one face of a tetramer that forms as a dimer of asymmetric dimers (Muraoka *et al*., 2003). Two DNA-binding domains act as a functional unit with local dyad symmetry constrained by the linker helices. The subunit orientation places the EBDs on one side of the tetramer such that their effector-mediated changes are additive. By the two-fold symmetry relationship of the tetramer, the two sets of DNA-binding domains are drawn together in response to effector binding. In the BenM model, each set of DNA-binding domains moves about 5 Å toward the center, creating a 10 Å total shift. The

direction of this movement is consistent with the changes in DNase I footprints that occur in the *benA* promoter region upon ligand binding, as described earlier and depicted in Figure 2.1 (Bundy *et al*., 2002, Ezezika *et al*., 2006).

Domain II appears to be an anchor point around which the tetramer is organized. Helix αH7 is at the interface of all the CatM- and BenM-EBD dimers. Helix αH6 of BenM is equivalent to a helix at the center of the CbnR tetramer, where two of the dimeric units associate. Thus, these helices from domain II are critical for high order oligomerization and orientation of the individual subunits. While domain II serves as an anchor, most effector-dependent movement is associated with domain I, which moves several Ångstroms toward domain II relative to its position in the absence of effectors (Figure 2.2). Significantly, helix αH3 moves 4-5 Å when muconate is bound. This helix abuts one DNA-binding domain and a linker helix that connects to another DNA-binding domain. Movement of this αH3 helix can impact the orientation of the DNA-binding domains and thus affect DNA binding. Other domain I regions move less but can also contribute to the overall contraction and the relative position of the subunits.

In contrast to the binding of muconate, the binding of benzoate to the primary site causes minimal conformational change in domain I. Despite having nearly identical primary effector-binding sites, CatM does not respond to benzoate as an inducer. Thus, BenM-activated transcription in response to benzoate alone may depend on the presence of benzoate in both the primary and secondary effector-binding sites. The affinity of benzoate for the secondary site is likely to be higher than that for the primary site, and the benzoate-protein interactions appear to be more specific in this latter site (Ezezika *et al*., 2007a). Benzoate and muconate may bind competitively in the primary effector-binding site, as suggested by the fluorescence emission spectroscopy studies (Clark *et al*., 2004b).

Our regulatory model predicts that maximal levels of *benA* expression result from a conformation of the DNA-BenM-effector-RNAP complex that can only be attained in the presence of both benzoate and muconate. Unfortunately, the structural characterization of any LTTR-DNA complex remains elusive. In the ligand-bound model of BenM, the inner DNA-binding domains are not sufficiently close to account for their apparent relative positions at Site 1 and Site 2 in the *benA* promoter region (Figure 2.1) (Bundy *et al*., 2002, Ezezika *et al*., 2006). It seems plausible that transcriptional activation by BenM involves local distortion of the DNA helix near *benA*. However, further work is needed to clarify how BenM and CatM interact with their target DNA and with RNAP.

Promoter specific regulation

One of the intriguing aspects of the BenM and CatM regulon is that the same regulator exerts dissimilar effects at different promoters. For example, the synergistic effect of benzoate and muconate on the BenM-activated transcription of *benA* is evident *in vivo* and *in vitro* (Bundy *et al*., 2002, Collier *et al*., 1998). However, such synergism has not been observed in the BenM-activated expression of the *benPK* operon (Clark *et al*., 2002, Collier *et al.*, 1997). At this locus, indicated in Figure 2.1B (diamond 1), BenM and CatM can each increase transcription in response to muconate but not to benzoate. Apparently, the conformational changes in BenM that occur upon binding benzoate do not lead to productive interactions with RNAP at the *benP* promoter as they do at the *benA* promoter.

The importance of the promoter sequence is further emphasized by amino acid substitutions in CatM whose effects on transcription depend on the target locus. An amino acid substitution in one variant, CatM(V158M), decreases muconate-inducible expression of the *catB*

promoter. In contrast, this same variant increases muconate-inducible expression of *benA* (Ezezika *et al*., 2006). Typically, CatM plays a minor role in expression of the *benABCDE* operon. As noted above, CatM recognizes and binds to the same *benA* promoter regions as BenM, most likely due to the similarity in their DNA-binding domains. However, muconate only enables CatM to activate low-level transcription from this promoter. CatM(V158M), unlike the wild-type CatM, activates *ben*-gene expression at high levels that are sufficient for BenM-independent growth on benzoate as the sole carbon source (Ezezika *et al*., 2006).

While protein-DNA interactions are not optimal for CatM-regulated *benA* expression, point mutations affecting either CatM or this promoter can significantly increase transcriptional activation. In a similar fashion to CatM(V158M), CatM(R156H) allows growth on benzoate without BenM by increasing CatM-regulated expression from the *benA* promoter (Ezezika *et al*., 2006). The CatM(R156H) variant was selected by its ability to activate inducer-independent expression of *catB* (Neidle *et al*., 1989). Interestingly, this variant has a different effect on *benA* and requires muconate for high-level transcription of this gene. When BenM and CatM are comparably positioned on the *benA* DNA, subtle structural differences may influence interactions with the transcriptional apparatus. At this locus, LTTR-DNA interactions are most favorable for control by BenM. Nevertheless, a single point mutation at -40, in the Site 2 region of *benA* (Figure 2.1C), increases CatM-dependent expression (Collier *et al*., 1998, Ezezika *et al*., 2006). Not only does this mutation enable wild-type CatM to substitute for BenM during growth on benzoate, but it also enables CatM(R156H) to activate *benA* transcription without muconate (Ezezika *et al*., 2006). Clearly, a full understanding of LTTR-mediated regulation requires analysis of the entire DNA-protein complex, including RNAP.

Mutational approaches to assess regulation

A facile selection method for regulatory mutants derives from the inability of BenM or CatM alone to permit wild-type growth rates on benzoate. Using parent strains that lack a functional copy of *benM* or *catM*, spontaneous mutants are readily isolated by growth on benzoate medium (Collier *et al*., 1998). Mutational analysis is assisted by the natural competence of strain ADP1. The high efficiency of DNA transformation and homologous recombination improves the ease with which chromosomal mutations can be localized and identified. Using this approach, mutations have been identified in regions expected to affect gene expression, such as in the *benA* promoter and in the *catM* gene (Collier *et al*., 1998, Ezezika *et al*., 2006).

This genetic method also has the power to reveal totally unpredicted regulatory effects. In one set of experiments, growth on benzoate yielded spontaneous mutants able to consume benzoate without either BenM or CatM (Reams & Neidle, 2003). These mutants compensate for the loss of the transcriptional regulators by increasing the copy number of *ben* and *cat* genes via gene duplication. In the selected mutants, gene amplification results in a tandem head-to-tail array of this genetic region on the chromosome. Multiple copies of weakly expressed genes can produce sufficient levels of the enzymes needed for benzoate catabolism. By exploiting the natural transformability of the *A. baylyi* strains, a novel experimental system was developed to study gene amplification, a common process with broad biological importance (Reams & Neidle, 2004a, Reams & Neidle, 2004b).

In a different set of experiments, mutations conferring the ability to grow on benzoate without BenM were localized in *catB*, the gene encoding muconate cycloisomerase (Cosper *et al.*, 2000). This result was unexpected since benzoate degradation requires a functional *catB* gene (Figure 2.1). In four independently isolated spontaneous mutants, a significant increase in CatM-

activated transcription from the *benA* promoter results from altered catalytic properties of a functional but variant muconate cycloisomerase enzyme. During growth on benzoate, the reduced activity of this enzyme, whose substrate is muconate, appears to increase intracellular muconate concentrations (Cosper *et al*., 2000). Thus, the ability of the wild-type CatM to activate *benA* transcription may be extremely sensitive to the intracellular concentration of its effector. Consistent with this model, the specific activity of catechol 1,2-dioxygenase (CatA), the enzyme that forms muconate, is elevated in cell extracts of the benzoate-grown *catB* mutants relative to the wild type. This elevation most likely results from the high muconate concentration causing an increase in CatM-dependent *catA* expression. Collectively, these results emphasize the importance of metabolite concentrations in this regulatory circuit.

These results also highlight the key role of transcription in modulating the endogenous accumulation of muconate, an effector that is toxic at high concentrations (Gaines *et al.*, 1996). In strains lacking *catB*, aromatic precursors such as benzoate, anthranilate, benzyl alcohol or benzaldehyde, are converted to muconate. In such strains, the toxic accumulation of this metabolite prevents growth on alternative carbon sources (Bundy *et al.*, 1998, Williams & Shaw, 1997). In contrast, exogenous muconate does not inhibit growth in the same fashion. It has been suggested that the coupled uptake and degradation of muconate prevents its intracellular accumulation and can account for the different effects of muconate generated endogenously or provided exogenously (Collier *et al*., 1997, Gaines *et al*., 1996, Brzostowicz *et al*., 2003).

Using parent strains that lack *catB*, direct selection yields spontaneous mutants that grow in the presence of a muconate-generating substrate (Bundy *et al*., 1998, Williams & Shaw, 1997). By this method, strains have been isolated in which growth can occur on a carbon source, such as succinate, because mutations block the conversion of the added substrate to muconate.

For example, numerous *catA* mutations have been obtained using this approach in which catechol accumulation is evident but not lethal. Thus, the endogenous accumulation of catechol appears to be less toxic than that of muconate. Nevertheless, the analysis of spontaneous gene amplification mutants suggests that in the presence of benzoate, expression of the *benABCDE* operon at levels much higher than that of *catA* leads to the catechol-dependent prevention of growth (Reams & Neidle, 2003, Reams & Neidle, 2004a). Thus, an important function of the BenM-CatM regulon is to regulate transcriptional levels for the balanced formation and degradation of metabolites to prevent the build up of toxic intermediates.

Cross regulation of 4-hydroxybenzoate catabolism by BenM and CatM

In studies of carbon source preferences, muconate was identified as a metabolite that inhibits 4-hydroxybenzoate (4HB) catabolism via the protocatechuate branch of the β–ketoadipate pathway (Figure 2.1) (Brzostowicz *et al*., 2003, Gaines *et al*., 1996). However, as noted above, the accumulation of high levels of muconate can inhibit the utilization of any carbon source. To determine if any effects of this compound were specific for 4HB degradation, studies were conducted of the *pob* and *pca* genes (Brzostowicz *et al*., 2003). Several lines of evidence, described next, suggest that muconate interacts with BenM and CatM to repress expression of the *pca* genes.

When *A. baylyi* ADP1 is fed a mixture of benzoate and 4HB, the consumption of 4HB is delayed until most of the benzoate has been fully degraded (Brzostowicz *et al*., 2003, Gaines *et al*., 1996). However, in mutants that can degrade benzoate without BenM or CatM, both compounds are consumed simultaneously. This observation suggests that BenM and CatM may repress the expression of genes used to degrade 4HB. A single functional copy of *catM* or *benM*

in benzoate-degrading mutants is sufficient for the preferential consumption of benzoate in the presence of 4HB (Brzostowicz *et al*., 2003). The need to eliminate both regulators to hasten 4HB utilization could be explained by their overlapping abilities to act as repressors. A similar overlap in the ability of BenM and CatM to activate the transcription of *catA* and *benP* requires the elimination of both regulators to abolish muconate-inducible expression of these loci (Clark *et al*., 2002, Romero-Arroyo *et al*., 1995).

Gel retardation assays were used to assess whether BenM and CatM bind DNA sequences that control 4HB degradation (Brzostowicz *et al*., 2003). Neither binds the promoter region of *pobA*, associated with the first step in 4HB catabolism (Figure 2.1). In contrast, both BenM and CatM bind specifically to a DNA fragment in the regulatory region of the divergently transcribed *pcaU* and *pcaI* genes. Consistent with muconate playing an inhibitory role, this compound enhances the binding of BenM or CatM to *pca*-region DNA (Brzostowicz *et al*., 2003). According to our current regulatory model, endogenous muconate formed during benzoate degradation enables BenM and CatM to bind this region and prevent PcaU from activating transcription from the *pcaI* promoter. This promoter controls a large operon that includes a key gene in the scheme, *pcaK* (D'Argenio *et al.*, 1999, Dal *et al.*, 2005). Our model proposes that the blocked induction of the PcaK transport protein sufficiently reduces or delays the uptake of 4HB into the cell to prevent the proper expression of *pobA*. As a result, the consumption of 4HB is inhibited during growth on benzoate.

Aspects of this model remain to be investigated further, as previously discussed (Brzostowicz *et al*., 2003). Nevertheless, ADP1 clearly controls the sequential order in which components in an aromatic compound mixture are consumed (Gaines *et al*., 1996). Such regulation may prevent mismatched interactions between the similarly structured enzymes and

substrates of the two branches of the β-ketoadipate pathway (Figure 2.1). Mismatched interactions must also be prevented between transcriptional regulators and metabolites that are chemically similar. Gene expression for the protocatechuate branch of ADP1 is activated by regulators of the IclR family, PobR and PcaU (Gerischer, 2002). This type of regulation differs from that in some bacteria where the expression of genes for protocatechuate degradation is regulated by a LTTR, PcaQ (Parke, 1996). In ADP1, proper regulation may rely on the distinct use of two regulatory protein classes. LTTRs and IclR regulators activate genes for the catechol and protocatchuate branch of the β-ketoadipate pathway, respective. As muconate-dependent repressors, BenM and CatM may help keep both branches from being simultaneously expressed at high levels.

Friends and Family: Additional *Acinetobacter* LTTRs

Homologs that appear to have similar functions to BenM and CatM are present in *A. lwoffii* K24, a bacterium that can use aniline (aminobenzene) as a sole source of carbon and nitrogen (Kim *et al.*, 2001). This compound is an abundant industrial chemical associated with plastics, pesticides, dyes and pharmaceuticals. The initial step in the aerobic catabolism of aniline is its conversion to catechol, which can subsequently be degraded by either the β–ketoadipate pathway or by a meta-cleavage pathway, depending on the microbe (Fujii *et al.*, 1997, Murakami *et al.*, 2003). Catechol degradation by *A. lwoffii* K24 and *A. baylyi* occurs via the β–ketoadipate pathway (Figure 2.1). Nevertheless, the genetic organization and regulation differ between the strains. In K24 there are two distinct sets of *cat* genes, each divergently transcribed from a putative LTTR gene (Kim *et al*., 2001). The genes for catechol dissimilation in one cluster are in the order *catBAC*. In the other cluster, the genes are in same order as those in

Pseudomonas putida, *catBCA* (McFall *et al*., 1998). The putative *A. lwoffii* LTTR regulators share approximately 55% sequence identity and are designated CatR1 and CatR2 based on their similarity to the corresponding LTTR of *P. putida.* Significant sequence identity is also observed between CatM of *A. baylyi* ADP1 and CatR1 (38%) or CatR2 (51%) of *A. lwoffii*. Gel retardation assays demonstrate that CatR1 can bind upstream of both *catB* genes. While the details remain to be clarified, CatR1 and CatR2 presumably regulate catechol degradation (Kim *et al*., 2001).

During aniline degradation by *A. lwoffii* K24, both sets of Cat enzymes are induced. However, the genes and enzymes needed for the conversion of aniline to catechol have not been identified in this strain. Typically, this conversion is mediated by an aniline dioxygenase, encoded by a well-conserved cluster of genes identified in bacteria of several genera, including *Acinetobacter* (Fujii *et al*., 1997, Murakami *et al*., 2003, Urata *et al.*, 2004). In the midst of this cluster is an open reading frame likely to encode another LTTR that regulates the conversion of aniline to catechol. Interestingly, aniline dioxygenase shares significant sequence similarity to benzoate dioxygenase (BenABC) whose genes are regulated by BenM in strain ADP1. Strain ADP1, which uses a different dioxygenase for the degradation of anthranilate (aminobenzoate), does not appear to encode an enzyme specific for aniline hydroxylation.

Clustered alongside the *ben* and *cat* genes on the ADP1 chromosome are genes for the catabolism of alkanoate esters (*are* genes) and alkyl salicylates (*sal* genes) (Jones *et al.*, 1999, Jones *et al*., 2000). These aromatic compounds are degraded through the β-ketoadipate pathway after conversion to either protocatechuate or catechol. Salicylate (2-hydroxybenzoate) is converted to catechol by salicylate hydroxylase, encoded by *salA*. As noted earlier, *salA* transcription is controlled by SalR (ACIAD1425; Table 1, no. 7). This LTTR is encoded immediately downstream of *salA* whose expression is induced by growth on salicylate or ethyl

salicylate. However, SalR does not regulate the *are* genes or *salDE*. The latter two genes appear to form an operon that is transcribed convergently relative to *salAR* (Jones *et al*., 1999, Jones *et al*., 2000).

Biosensors, biotechnology, and other applications for LTTRs

A recently developed method for monitoring salicylate depends on the SalR-regulated expression of chromosomal transcriptional fusions in ADP1-derived strains (Huang *et al.*, 2006, Huang *et al.*, 2005). Salicylate is produced as a metabolite during the aerobic microbial degradation of aromatic pollutants such as naphthalene, phenanthrene and anthracene. Therefore, a sensitive detection system for salicylate might be helpful in assessing the level of pollutants and microbial activities in samples from contaminated environments. Additional reasons to monitor salicylate levels derive from its potential to stimulate the degradation of some pollutants and its role in plants as an inducer of systemic acquired resistance to pathogens (Singer *et al.*, 2003).

A reporter gene inserted into the ADP1 chromosome between *salA* and *salR* allows a correlation to be made between the levels of the reporter-encoded protein(s) and the salicylate-induced activation of transcription by SalR. Two types of reporters were successfully used for this purpose, one encoding green fluorescent protein and the other using *luxCDABE*-encoded bioluminescence (Huang *et al*., 2005). The utility of this whole-cell bioreporter system was confirmed in experiments to monitor salicylate production by a culture of naphthalene-degrading *Pseudomonas putida* (Huang *et al*., 2005). Furthermore, the *Acinetobacter* biosensor method was successful in detecting salicylate *in situ* within tobacco leaves (Huang *et al*., 2006).

Several review articles discuss the development of whole-cell biosensors, their potential applications, and features that may limit their utility beyond academic studies (Daunert *et al.*, 2000, Harms *et al.*, 2006, van der Meer *et al.*, 2004). One concern with the use of modified bacteria is the possible release of genetically engineered plasmids in the environment. However, the integration of a reporter gene into a bacterial chromosome can improve the stable containment of modified DNA. ADP1 affords advantages for biosensor development, as its chromosome can be specifically manipulated with ease in the laboratory.

Additional problems with biosensor methodology may arise if the regulatory protein that controls expression of the reporter gene has a low affinity for its effector(s) or if the specificity of the regulator does not permit the desired compound to induce gene expression. SalR, NahR and other LTTRs can be used as biosensors to detect aromatic pollutants (Daunert *et al*., 2000, van der Meer *et al*., 2004). Improved understanding of the effector-regulator interactions controlling aromatic compound degradation is emerging from the structural analysis of LTTRs such as CbnR (Muraoka *et al*., 2003), DntR (Smirnova *et al*., 2004), BenM, and CatM (Ezezika *et al*., 2007a). Mutational approaches should facilitate the isolation of variant regulators with properties needed for specific applications. For example, targeted mutation may be used to develop DntR as a biosensor for nitrotoluenes (Smirnova *et al*., 2004).

Most LTTRs studied in *Acinetobacter* strains control the degradation of aromatic compounds. Applications for these regulators include the generation of desired chemicals through biotransformation. For example, *cis*-diols formed during the metabolism of aromatic compounds can serve as chiral synthons to produce useful compounds such as pharmaceuticals (Raschke *et al.*, 2001). Since LTTRs often regulate the catalytic steps that form these metabolites, regulatory manipulation may improve the bioconversion yield. A successful strategy

to engineer chromosomal genes in ADP1 recently involved *argR* (Elbahloul & Steinbuchel, 2006). Although this gene encodes a regulator in the AsnC family, similar methods should work for LTTRs. In the *argR* example, several metabolic steps were manipulated to enhance the biosynthesis of cyanophycin, which can be used in chemical and medical applications. Medical applications for LTTRs are likely to expand. For instance, LTTRs might serve as targets for drugs that are selectively toxic to bacteria. While such applications have yet to be developed, LTTRs are medically relevant and can contribute to virulence. As the focus on multidrug resistant *A. baumannii* intensifies, new LTTRs continue to be identified that appear to regulate antibiotic resistance (Fournier *et al*., 2006, Mussi *et al.*, 2005). Further studies are needed to understand the role of these and other LTTRs in virulence and drug resistance in pathogenic *Acinetobacter* species.

Conclusions

Since the LysR-type regulators were first recognized as a family, the number of known and predicted LTTRs has grown enormously. Considering the diversity of LTTR-regulated functions, the wide range of organisms that encode them, and the variations among member sequences, it is remarkable that so many common features are evident. These regulators are characterized by a wHTH motif in the N-terminal DNA-binding domain followed by a relatively long helix that connects to a two-domain region for effector recognition and response (Muraoka *et al*., 2003, Smirnova *et al*., 2004). The structure of this latter region is essentially that of a periplasmic binding protein in which a ligand-binding pocket forms between the two domains (Quiocho & Ledvina, 1996). While this general structure appears to be a shared aspect of LTTRs, differences in regulation by individual proteins provide illuminating comparisons. The

CatM regulator of *A. baylyi* ADP1, one of the earliest known family members, serves as a model for understanding LTTR interactions with effectors and promoters (Ezezika *et al*., 2006, Neidle *et al*., 1989, Romero-Arroyo *et al*., 1995). Comparative studies of CatM and a paralog, BenM, reveal that two distinct effector-binding sites in BenM underlie its ability to activate transcription synergistically in response to benzoate and muconate (Bundy *et al*., 2002, Ezezika *et al*., 2007a). As genetic studies of the BenM-CatM regulon illustrate, the natural transformation system of ADP1 makes *A. baylyi* ideal for laboratory investigations of complex regulation. Approximately 25% of the transcriptional regulators in ADP1 are predicted to be LTTRs. Our initial analysis of these family members provides a starting point for studies of their function. As LTTRs in other *Acinetobacter* strains and species are also explored further, an improved understanding should emerge concerning genetic regulation, metabolism, pathogenicity, and drug resistance.

Among the greatest remaining challenges in understanding LTTRs is to elucidate key interactions within the entire protein-DNA complex. Currently, there are no known structures of LTTRs bound to DNA in either active or inactive conformations. While many regulatory mutants have revealed the consequences of altered sequences in LTTR proteins and their target DNA, precise interpretation of the data requires more detailed knowledge of atomic-level structure. Specific contacts with RNAP that are involved in LTTR-mediated activation are not yet well understood. Nevertheless, these limitations are balanced by a renewed enthusiasm for regulatory investigations that stem from rapid increases in the availability of genomic sequences. Furthermore, as microarray techniques are applied to more organisms, appreciation for the importance of LTTRs is growing. In many cases, LTTRs are at the heart of complex and physiologically important regulatory circuits. Recent discoveries are stimulating new

experiments on LTTR structure and function that are helping to meet the goals of basic and applied research.

CHAPTER 3

INDUCER RESPONSES OF BENM, A LYSR-TYPE TRANSCRIPTIONAL REGULATOR IN *ACINETOBACTER BAYLYI* ADP1[2]

[2] Craven, S. H., O. C. Ezezika, S. Haddad, R. A. Hall, C. Momany, and E. L. Neidle. 2009. *Molecular Microbiology*. **72**[4], 881-894. Reprinted here with permission of publisher.

Summary and Introduction

BenM and CatM control transcription of a complex regulon for aromatic compound degradation. These *Acinetobacter baylyi* paralogs belong to the largest family of prokaryotic transcriptional regulators, the LysR-type proteins. Whereas BenM activates transcription synergistically in response to two effectors, benzoate and *cis,cis*-muconate, CatM responds only to *cis,cis*-muconate. Here, site-directed mutagenesis was used to determine the physiological significance of an unexpected benzoate-binding pocket in BenM discovered during structural studies. Residues in BenM were changed to match those of CatM in this hydrophobic pocket. Two BenM residues, R160 and Y293, were found to mediate the response to benzoate. Additionally, alteration of these residues caused benzoate to inhibit activation by *cis,cis*-muconate, positioned in a separate primary effector-binding site of BenM. The location of the primary site, in an inter-domain cleft, is conserved in diverse LysR-type regulators. To improve understanding of this important family, additional regulatory mutants were analyzed. The atomic-level structures were characterized of the effector-binding domains of variants that do not require inducers for activation, CatM(R156H) and BenM(R156H,T157S). These structures clearly resemble those of the wild-type proteins in their activated muconate-bound complexes. Amino acid replacements that enable activation without effectors reside at protein interfaces that may impact transcription through effects on oligomerization.

Introduction

Nearly 20% of all transcriptional regulators identified by bacterial genome analysis are LysR-type transcriptional regulators (LTTRs), the largest regulatory family in prokaryotes (Pareja *et al*., 2006). Despite this prevalence, the molecular basis of effector-induced activation

by LTTRs remains unclear. Aggregation problems impede structural studies, and there are only two known atomic-level structures of LTTRs bound to their cognate inducers (Ezezika *et al.*, 2007a). These structures encompass the effector-binding domains (EBDs) of CatM and BenM, two similar LTTRs that control benzoate degradation in the soil bacterium *Acinetobacter baylyi* ADP1 (Figure 3.1) (Collier *et al*., 1998, Romero-Arroyo *et al.*, 1995). BenM- and CatM-EBD both bind an inducer in an interdomain cleft. Surprisingly, a secondary effector-binding site was discovered in a hydrophobic pocket of BenM-EBD. In the current investigation, the physiological significance of this discovery was tested. LTTR-inducer interactions were explored further by characterizing mutants with atypical responses to effectors.

In LTTRs, the N-terminal DNA-binding domain (DBD) is linked to the EBD, which extends to the C-terminus and is composed of two subdomains. The characterized structures of one full-length LTTR and the regulatory regions of several others demonstrate that the protein fold of the EBD is essentially that of a periplasmic-binding protein (Choi *et al*., 2001, Muraoka *et al*., 2003, Quiocho & Ledvina, 1996, Smirnova *et al*., 2004, Stec *et al.*, 2006). The structural conservation in this domain is remarkable in light of the sequence variability in different LTTRs. For example, the structurally similar EBDs of BenM and CysB share only 12% sequence identity (Tyrrell *et al*., 1997, Verschueren *et al*., 1999). Consistent with the typical role of the EBD in binding a low molecular weight inducer, LTTRs that respond to the same or similar compounds share a higher degree of sequence similarity.

CatM and BenM, which are 59% identical in sequence, both respond to the metabolite *cis,cis*-muconate (hereafter called muconate; Figure 3.1) (Collier *et al*., 1998, Romero-Arroyo *et al*., 1995). These two LTTRs have overlapping roles in a complex regulon for aromatic

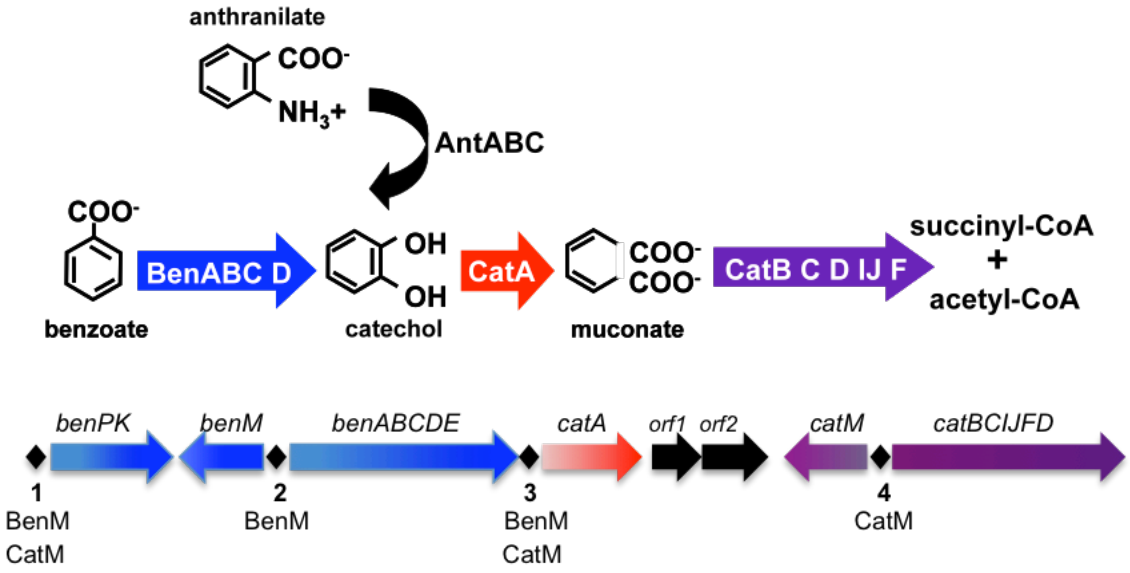


Figure 3.1: Pathway for the degradation of benzoate and anthranilate. LysR-type transcriptional regulators, BenM and CatM, control expression from multiple promoters (labeled 1-4) in a supraoperonic cluster of chromosomal genes. At two loci (1 and 3), BenM and CatM play equally important regulatory roles. At two loci (2 and 4), one regulator is primarily responsible for gene expression (BenM or CatM, respectively). The *benPK* operon encodes for transport proteins (Clark *et al*., 2002). The *antABC* genes reside on a distant region of the chromosome and are not regulated by BenM or CatM (Bundy *et al*., 1998).

compound degradation. A distinguishing characteristic of BenM is its additional ability to recognize benzoate as an inducer. At the promoter of the *benABCDE* operon, BenM activates gene expression in response to either benzoate or muconate. When present together, the two compounds have a synergistic effect on transcriptional activation (Bundy *et al*., 2002). This unusual type of synergism enables the bacterium to integrate multiple metabolic signals rapidly and to prevent the buildup of toxic intermediates during benzoate consumption.

Structural studies raised the possibility that synergism depends on conformational changes that occur when muconate binds in the primary effector-binding site and benzoate binds in the secondary site (Figure 3.2A). However, the studies of effector-soaked BenM-EBD crystals involved high, non-physiological concentrations of benzoate (120 mM) (Ezezika *et al*., 2007a). Therefore, it was important to test whether the newly discovered secondary effector-binding site is critical for benzoate-dependent transcription. In the BenM-EBD structure, a benzoate molecule in the secondary site contacts an arginine and a tyrosine at positions 160 and 293 (Figure 3.2B). Although CatM does not respond to benzoate, its amino acids corresponding to this pocket are substantially conserved in BenM with the exception of residues 160 and 293. As described herein, these amino acids in BenM were altered. The effects of specific *benM* mutations on growth and gene expression were tested. In cases where the resulting mutants failed to consume benzoate, it was possible to select and analyze spontaneous mutations that restored the ability to grow with benzoate as the sole carbon source.

Furthermore, arginine 156 of BenM was replaced with histidine. This variant was generated to resemble CatM(R156H), a regulator that activates transcription from the *catA* and *catB* promoters in the absence of an inducer (Ezezika *et al.*, 2006, Neidle *et al.*, 1989). This

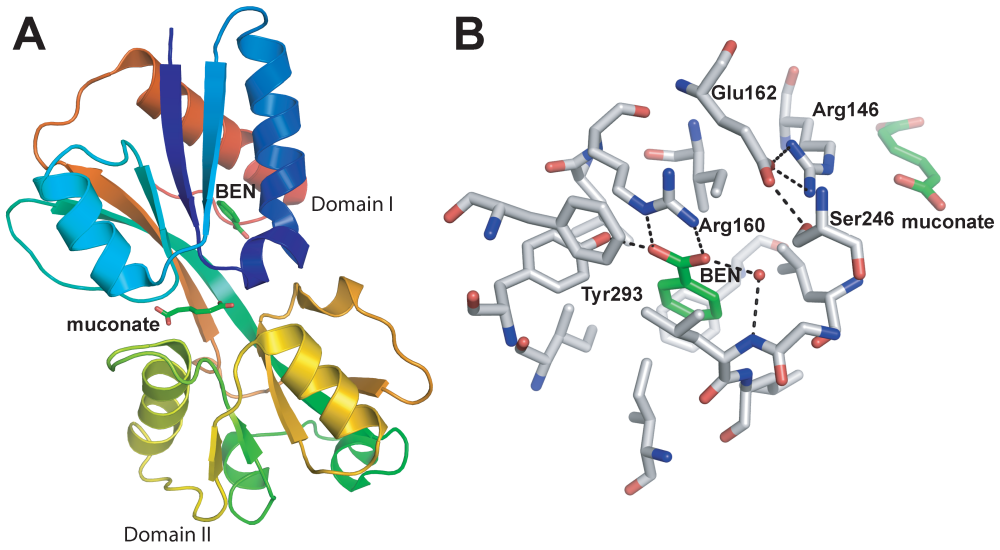


Figure 3.2: Representations of the BenM-EBD structure. A BenM-EBD subunit, in ribbon representation, shows the inducers muconate and benzoate (BEN) bound in the primary and secondary effector-binding sites respectively (A). In the region of the secondary effector-binding site, Arg160 andTyr293 make direct contact with a bound benzoate molecule (B). A model for electrostatic control of synergism entails communication of the binding of benzoate via interactions among Arg160, Glu162 and Arg146, a residue near muconate (Ezezika *et al*., 2007a). Residues 160 and 293 differ in CatM, which has histidine and phenylalanine at these sites, respectively.

replacement, in the central portion of the EBD, is not directly adjacent to either of the structurally identified effector-binding sites. To help determine the role of this residue in mediating the response to inducers, protein structures were characterized by X-ray crystallographic methods. Here we report the structures of CatM(R156H)-EBD and BenM(R156H,T157S)-EBD and discuss conformational changes involved in transcriptional activation by LTTRs.

Results

Testing the role of R160 and Y293 of BenM in benzoate-dependent transcriptional activation

Replacements were engineered in two BenM residues predicted to mediate benzoate binding (Figure 3.2B). Plasmid pBAC713 (Table 3.1), encoding BenM(R160H,Y293F), generates a secondary effector-binding region that should resemble that of CatM, which does not respond to benzoate. DNA from pBAC713 replaced the wild-type chromosomal *benM* via allelic exchange to generate strain ACN641 (Table 3.1). ACN641 was able to grow with muconate or anthranilate as a sole carbon source (Figure 3.1). Thus, the *cat*-genes can be expressed despite the *benM* mutations. In contrast, ACN641 did not consume benzoate. To test whether lack of growth stemmed from low *ben*-gene expression, a *benA*::*lacZ* transcriptional fusion was used to replace *benA* on the ACN641 chromosome, thereby generating ACN661 (Table 3.1). LacZ activity was measured in ACN661 and ACN32, a comparable strain with wild-type *benM* (Figure 3.3).

Table 3.1: Bacterial strains and plasmids.

Strain	Relevant Characteristics[a]	Reference or Source
***A. baylyi* strains**		
ADP1	wild-type (BD413)	Juni and Janik, 1969
ACN32	*benA*::*lacZ*-Km^r*5032*	Collier *et al.*, 1998
ACN47	*benM*::ΩS*4036 benA*::*lacZ*-Km^r*5032*	Collier *et al.*, 1998
ACN532	*benM5532* [BenM(R156H,T157S)] *benA*::*lacZ*-Km^r*5032* Δ*catMBC3205*	This Study
ACN533	*benM5532* [BenM(R156H,T157S)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN637	*benM*::*sacB*-Km^r*5624*	This Study
ACN639	*benM5639* [BenM(R160M,Y293F)]	This Study
ACN640	*benM5640* [BenM(R160K,Y293F)]	This Study
ACN641	*benM5641* [BenM(R160H,Y293F)]	This Study
ACN659	*benM5639* [BenM(R160M,Y293F)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN660	*benM5640* [BenM(R160K,Y293F)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN661	*benM5641* [BenM(R160H,Y293F)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN678	*benM*5678 [BenM(R156H)]	This Study
ACN680	*benM*5678 [BenM(R156H)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN795	*benM5795* [BenM(R160H)]	This Study

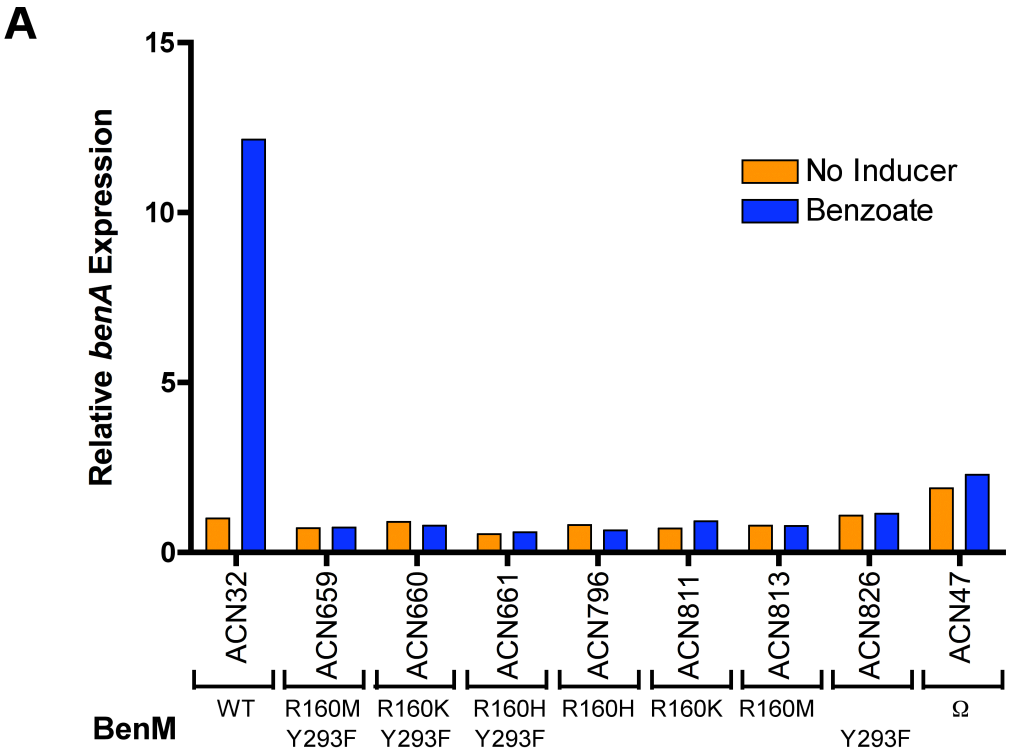
Strain	Relevant Characteristics[a]	Reference or Source
ACN796	*benM5795* [BenM(R160H)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN807	*benM5807* [BenM(R160K,R225H,Y293F)]:spontaneous ben+ mutant of ACN640	This Study
ACN809	*benM5809* [BenM(R160H,E226K,Y293F)]: spontaneous ben+ mutant of ACN641	This Study
ACN810	*benM5810* [BenM(R160K)]	This Study
ACN811	*benM5810* [BenM(R160K)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN812	*benM5812* [BenM(R160M)]	This Study
ACN813	*benM5812* [BenM(R160M)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN824	*benM5824* [BenM(Y293F)]	This Study
ACN826	*benM5824* [BenM(Y293F)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN828	*benM5807* [BenM(R160K,R225H,Y293F)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN829	*benM5809* BenM[(R160H,E226K,Y293F)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN864	*benM5864* [BenM(R225H)]	This Study
ACN865	*benM5864* [BenM(R225H)] *benA*::*lacZ*-Km^r*5032*	This Study
ACN866	*benM5866* [BenM(E226K)]	This Study
ACN867	*benM5866* [BenM(E226K)] *benA*::*lacZ*-Km^r*5032*	This Study
XL1_Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q}ZΔ*M15* Tn*10* (Tet^r)]	Stratagene

Plasmids		
pUC19	Ap^r; cloning vector	Invitrogen
pET-21b	Ap^r; T7 expression vector	Novagen
pCR2.1-TOPO	Ap^rKm^r; PCR cloning vector	Invitrogen
pBAC7	Ap^r; *benKM* (563-2964) [b] region in pUC19	This Study
pIB17	Ap^r; *catM3102* (11950-13205) [b] in pUC19	Neidle *et al.*, 1989
pBAC54	Ap^rKm^r; *lacZ*-Km^r cassette in NsiI site (3761) [b] in *benA* (2316-5663) [b] in pUC19	Collier *et al.*, 1998
pBAC430	Ap^r; *catM* (12116-13027) [b] in pET-21b	Clark *et al.*, 2004a
pBAC433	Ap^r; *benM* (1453-2368) [b] in pET-21b	Clark *et al.*, 2004a
pBAC668	Ap^r; *benKM* (563-2964) [b] mutant allele encodes [BenM(R156H,T157S)]	This Study
pBAC680	Ap^r Km^r; PCR fragment of *catM3102* (12119-12787) [b] from PIB17 in pCR2.1-TOPO: [CatM(R156H)]	This Study
pBAC683	Ap^r; *NdeI-XhoI* (12119-12787) [b] fragment from pBAC680 in pET-21b: Expression construct for [CatM(R156H)]-EBD	This Study
pBAC692	Ap^r Km^r; PCR fragment of *benM5532*(1457-2368) [b] in pCR2.1-TOPO: [BenM(R156H,T157S)]	This Study
pBAC698	Ap^r; *NdeI-XhoI* (1457-2128) [b] fragment from pBAC692 in pET-21b: Expression construct for [BenM(R156H,T157S)]-EBD	This Study
pBAC709	Ap^r Km^r; *benKM* (563-2316) [b] pUC19. Contains *sacB*::*kan* counterselectable marker in *Sal*I site (1930) [b] of *benM* gene	This Study
pBAC711	Ap^r; *benM* (1453-2368) [b] mutant allele encodes [BenM(R160M,Y293F)]	This Study

pBAC712	Ap^r; *benKM* (563-2964) [b] mutant allele encodes [BenM(R160K,Y293F)]	This Study
pBAC713	Ap^r; *benKM* (563-2964) [b] mutant allele encodes [BenM(R160H,Y293F)]	This Study
pBAC714	Ap^r; *benM* (1453-2368) [b] mutant allele encodes [BenM(R160H,Y293F)]	This study
pBAC769	Ap^r; *benM* (1453-2368) [b] mutant allele encodes [BenM(R160H)]	This Study
pBAC770	Ap^r; *benM* (1453-2368) [b] mutant allele encodes [BenM(R160K)]	This Study
pBAC771	Ap^r; *benM* (1453-2368) [b] mutant allele encodes [BenM(R160M)]	This Study
pBAC772	Ap^r; *benM* (1453-2368) [b] mutant allele encodes [BenM(Y293F)]	This Study
pBAC776	Ap^r; *benM* (1453-2368) [b] mutant allele encodes [BenM(R225H)]	This Study
pBAC778	Ap^r; *benM* (1453-2368) [b] mutant allele encodes [BenM(E226K)]	This Study

[a]Ap^r, ampicillin resistant; Sm^r, streptomycin resistant; Sp^r, spectinomycin resistant; Km^r, kanamycin resistant; Ω, omega cassette conferring Sm^rSp^r (Prenti and Kirsch, 1984); *sacB*, counterselectable marker, used as described previously (Jones and Williams, 2003).

[b]Position in the *ben-cat* sequence in GenBank entry AF009224.


A
Relative benA Expression
15
10
5
0
No Inducer
Benzoate
ACN32
ACN659
ACN660
ACN661
ACN796
ACN811
ACN813
ACN826
ACN47
BenM
WT
R160M Y293F
R160K Y293F
R160H Y293F
R160H
R160K
R160M
Y293F
Ω

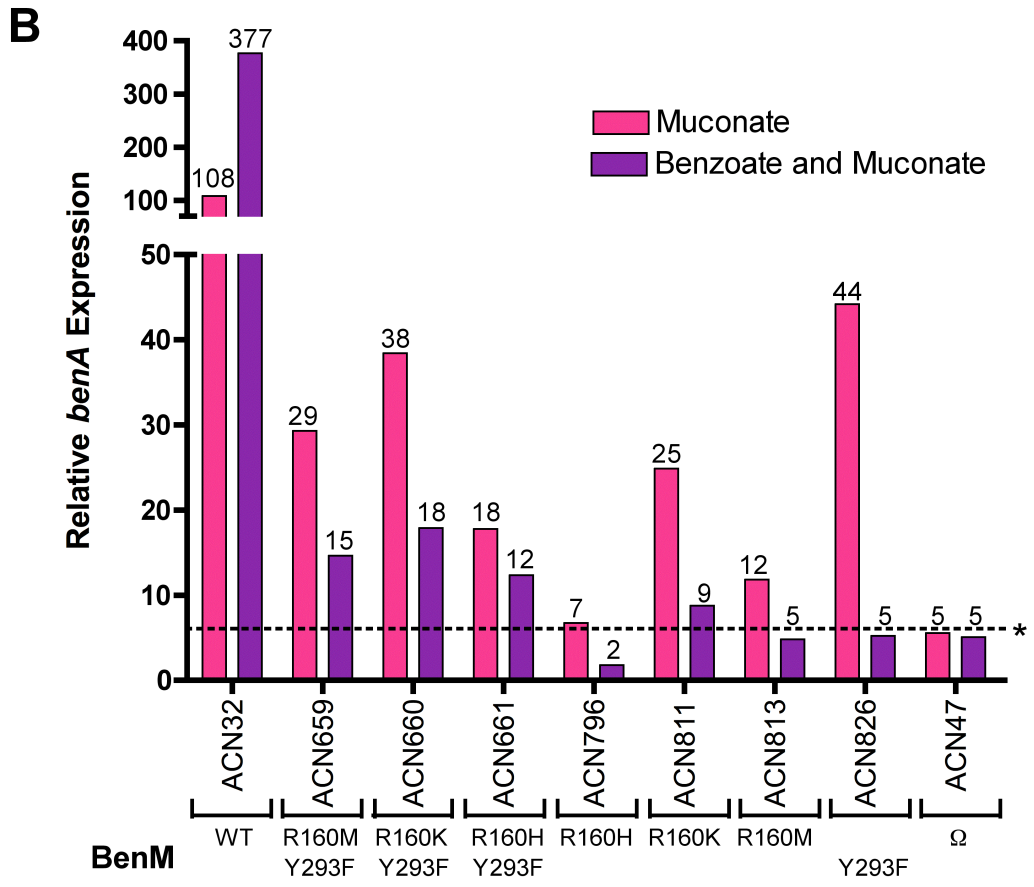

B
Relative benA Expression
400
300
200
100
50
40
30
20
10
0
Muconate
Benzoate and Muconate
377
108
29
15
38
18
18
12
7
2
25
9
12
5
44
5
5
5
*
ACN32
ACN659
ACN660
ACN661
ACN796
ACN811
ACN813
ACN826
ACN47
BenM
WT
R160M Y293F
R160K Y293F
R160H Y293F
R160H
R160K
R160M
Y293F
Ω

Figure 3.3: Effect of *benM* alleles on expression of a chromosomal *benA*::*lacZ* transcriptional reporter. Strains and their encoded BenM variants are indicated. Controls include ACN32, with wild-type *benM* (WT), and ACN47, with an inactivated *benM* (Ω). Relative *benA* expression is reported as the ratio of measured LacZ activity to that in uninduced ACN32, 2.90 ± 0.37 (nmol/min/ml/OD_{600}). Benzoate (1mM) was provided as an inducer in **(A)**. Muconate (1 mM) or both compounds (0.5 mM benzoate, 0.5 mM muconate) were also tested as inducers in **(B)**. The dotted line (*) represents the level of CatM-dependent *benA* expression in the presence of muconate, as observed in ACN47. All strains encode wild-type CatM. Data represent the average of four to eight replicates, and standard deviations were within 15% of the average value.

The chromosomal disruption of *benA* prevents further metabolism of benzoate and allows its effects as an inducer to be examined. Whereas the addition of benzoate induced a 12-fold increase in gene expression in ACN32, benzoate-dependent *benA* expression was abolished in ACN661 (Figure 3.3A). In contrast, muconate increased *benA*::*lacZ* expression 18 fold in ACN661 (Figure 3.3B). Although this muconate-inducible expression is lower than with wild-type BenM (ACN32), it remains higher than in the absence of BenM (ACN47) (Figure 3.3B). In the latter strain, CatM mediates a low level of muconate-inducible *benA* expression (Collier *et al*., 1998). Collectively, these results indicate that BenM(R160H,F293Y) activates *benA* expression in response to muconate but not benzoate. Without activation in response to benzoate, it is not surprising that the synergistic effect of both inducers, seen in ACN32, was also lost in ACN661. Interestingly, when both effectors were added to ACN661, benzoate inhibited muconate-activated *benA* expression (Figure 3.3B).

Individual amino acid replacements at positions 160 and 293 in BenM

Site-directed mutagenesis was used to create BenM variants with either R160H or Y293F. Using allelic exchange methods, *benM* on the chromosome was replaced with plasmid-borne alleles. The individual amino acid replacements had similar effects to the double replacement. ACN795 and ACN824 encoding BenM(R160H) or BenM(Y293F), did not grow on benzoate. Moreover, there was no benzoate-activated *benA* expression in strains with a chromosomal *lacZ* transcriptional fusion (ACN796 and ACN826, Figure 3.3A). In ACN826, the muconate-induced *benA* expression, mediated by BenM(Y293F), was strongly inhibited by benzoate (Figure 3.3B). Since muconate-induced *benA*

expression in ACN796 was not significantly higher than in the absence of BenM (in ACN47), the BenM(R160H) protein may not be stably produced. Nevertheless, the inhibition of muconate-activated expression in ACN796 but not ACN47 may indicate that *benA* in the former strain is regulated by BenM(R160H), albeit at a low level.

To investigate position 160 in BenM further, arginine was replaced with lysine or methionine, amino acids of similar charge or size to the wild-type residue. The resulting strains, ACN810 and ACN812, which encode BenM(R160K) or BenM(R160M), respectively, failed to grow on benzoate. In the corresponding strains with a chromosomal *benA*::*lacZ* reporter (ACN811 and ACN813), gene expression indicates that both BenM variants activate transcription in response to muconate but not benzoate (Figure 3.3B). However, with either replacement at residue 160, muconate-activated *benA* expression was inhibited by benzoate.

BenM(R160M,Y293F) enables growth on benzoate as a sole carbon and energy source

Additional variants were made to combine each replacement at position 160 with Y293F. When these *benM* alleles controlled *benA*::*lacZ* expression in ACN659 and ACN660, the double variants, BenM(R160M,Y293F) and BenM(R160K,Y293F), showed similar expression patterns to their single replacement counterparts (Figure 3.3). However, unlike the other mutants, ACN639, encoding BenM(R160M,Y293F), grew on benzoate as the carbon source, although more slowly and with a longer lag time than the wild type (Table 3.2). To assess the possibility that ACN639 had acquired additional mutations, plasmid-derived *benM* DNA was used to transform a recipient encoding BenM(R160M) or BenM(Y293F). Growth on benzoate occurred only when

Table 3.2: Effect of *benM* mutations on growth with benzoate as the sole carbon source.[a]

Strain	Relevant characteristics	Generation time (min)[b]	Lag time (min)[c]
ADP1	Wild type	103 ± 5	95 ± 10
ACN639	BenM(R160M,Y293F)	147 ± 26	200 ± 30
ACN678	BenM(R156H)	83 ± 1	60 ± 5
ACN866	BenM(E226K)	85 ± 5	65 ± 5

[a]All strains had comparable growth rates with succinate as the sole carbon source (data not shown).
[b]Averages of at least four determinations.
[c]Time between inoculation and start of exponential growth.

recombination could generate an allele encoding BenM(R160M,Y293F), data not shown. The transformation rates suggested that additional mutations are not needed for growth on benzoate.

Since the BenM(R160M,Y293F) variant does not respond to benzoate (ACN659, Figure 3.3A), it was not obvious how ACN639 expresses *benABCDE* sufficiently for benzoate consumption. In some mutants lacking BenM, increased levels of muconate-inducible *benA* expression permit growth on benzoate (Cosper *et al*., 2000, Ezezika *et al*., 2006). However, the muconate-dependent expression mediated by BenM(R160M,Y293F) was comparable to that of other variants that do not confer growth on benzoate (Figure 3.3B). Since these assays used high concentrations of extracellular effectors, conditions were modified to mimic wild-type metabolism more closely. The expression of *benA*::*lacZ* was monitored during growth on anthranilate, a substrate that generates intracellular muconate during catabolism (Figure 3.1). Under these conditions, BenM(R160M,Y293F) yielded higher *benA* expression than wild-type BenM or the variants with individual replacements (Figure 3.4). This relatively high expression was most pronounced at early stages in the growth cycle of batch cultures (OD_{600}~0.03-0.07). Thus, the early activation of *benA* expression by BenM(R160M,Y293F) may enable ACN639 to grow on benzoate. This variant may respond to relatively low intracellular concentrations of muconate.

Selection of spontaneous mutants that grow on benzoate (ben$^+$phenotype)

Mutants ACN640 and ACN641, which do not grow on benzoate, gave rise to spontaneous colonies on benzoate medium. Two such ben$^+$ derivatives, ACN807 and

ACN809, were characterized. DNA sequencing confirmed that ACN807 retained the mutations of its parent (ACN640) encoding R160K and Y293F. Similarly, ACN809 retained the mutations of its parent (ACN641) encoding R160H and Y293F. Each strain also had a *benM* mutation encoding a third amino acid change, R225H in ACN807 and E226K in ACN809. Gene expression was assessed in strains with a chromosomal *benA*::*lacZ* fusion controlled by an adjacent *benM* allele encoding either BenM(R160K,R225H,Y293F) (ACN828) or BenM(R160H,E226K,Y293F) (ACN829). In both cases, there were high levels of *benA* expression, with or without the addition of muconate and/or benzoate as inducers (Figure 3.5). The R225H and E226K replacements each enabled higher *benA* expression in the absence of inducer than did muconate-induced wild-type BenM.

Transcriptional activation by BenM without inducers

To determine the individual effects of R225H or E226K, engineered *benM* alleles on plasmids pBAC776 and pBAC778 were substituted for the chromosomal *benM*. Strains ACN864, encoding BenM(R225H), and ACN866, encoding BenM(E226K), grew on benzoate as the carbon source (Table 3.2). This ben$^+$ phenotype was consistent with having wild-type sequences in the secondary effector-binding site of BenM. Reporter strains subsequently demonstrated that the replacements at residue 225 or 226 affected gene expression. In ACN865 and ACN867, an allele encoding BenM(R225H) or

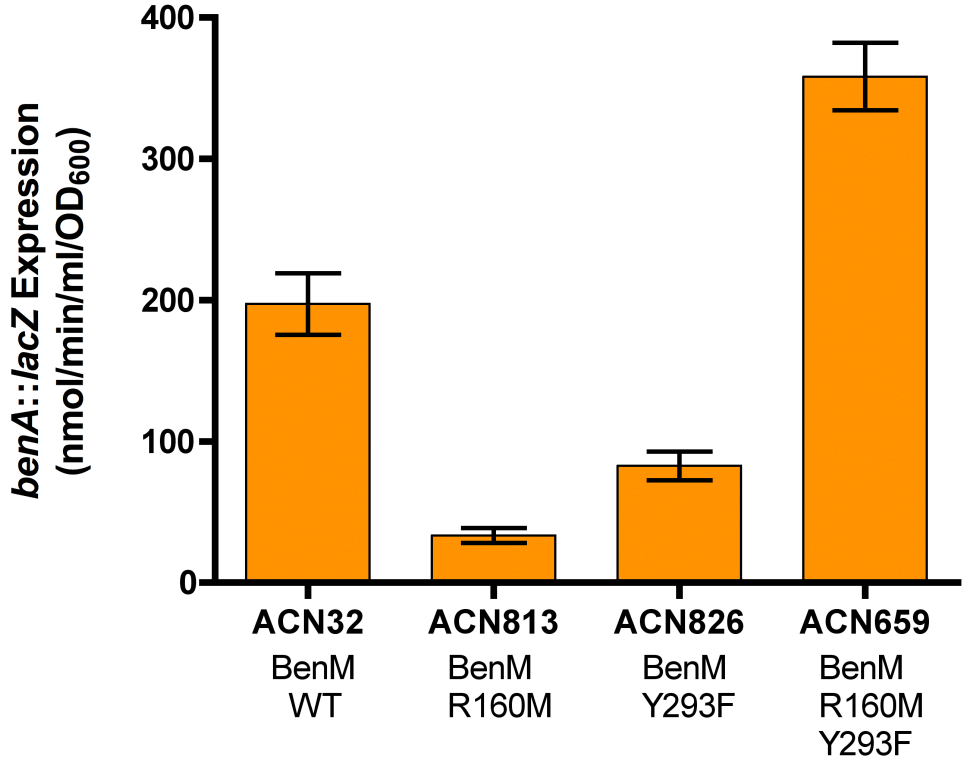


Figure 3.4: Effect of *benM* alleles on *benA*::*lacZ* expression in anthranilate-grown cultures. LacZ activity was assayed early in the growth phase (OD_{600} ~ 0.03-0.04) for the indicated strains provided with 1 mM anthranilate as the sole carbon source. ACN32 encodes a wild-type BenM (WT), and the other strains encode variants (as shown). Data represent the average of five replicates with standard deviations shown by error bars.

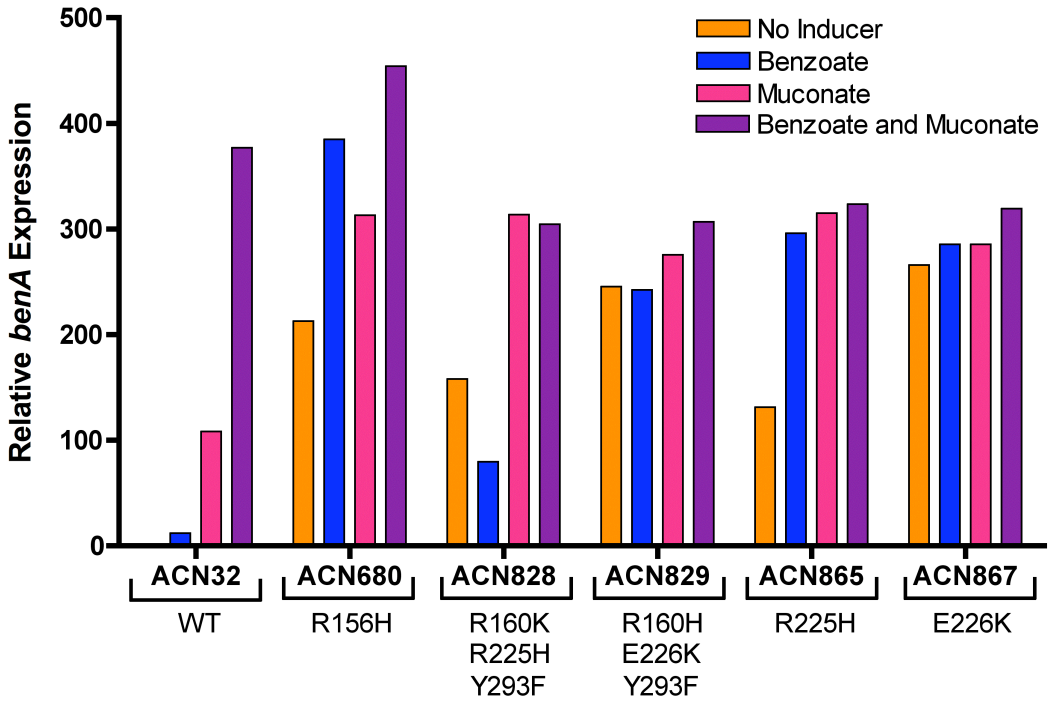


Figure 3.5: Expression of a chromosomal *benA*::*lacZ* fusion in strains encoding BenM or BenM variants. Cultures were grown in LB with addition of the indicated inducer (1 mM benzoate, 1 mM muconate, or 0.5 mM benzoate and 0.5 mM muconate). LacZ activity is reported relative to uninduced ACN32 (2.90 ± 0.37 nmol/min/ml/OD_{600}). Activities are the average of at least four repetitions with standard deviations within 20% of the average value.

BenM(E226K), respectively, controlled an adjacent chromosomal *benA*::*lacZ* fusion. Both variants activated high-level *benA* expression without added inducers (Figure 3.5). The uninduced expression of *benA* in ACN867 was 266-fold higher than with wild-type BenM and was not substantially increased by benzoate or muconate. In ACN865, encoding R225H with a wild-type secondary effector-binding site, benzoate alone activated *benA* expression. In contrast, the same R225H replacement in combination with R160K and Y293F in ACN828 caused inhibition of *benA* expression when benzoate was provided as the only effector.

To improve understanding of transcriptional activation without inducers, a variant was constructed to resemble a previously studied homolog, CatM(R156H), known to induce CatA and CatB without muconate (Neidle *et al*., 1989). ACN678 was created with a chromosomal allele that encodes BenM(R156H). This strain grew well with benzoate as the carbon source (Table 2). Another strain, ACN680, was generated in which this allele controlled expression of the *benA*::*lacZ* transcriptional fusion. In the absence of inducer, *benA* expression in ACN680 under the control of BenM(R156H) was 213-fold higher than in ACN32 with wild-type BenM (Figure 3.5). The addition of inducers to ACN680 elicited a further increase in gene expression.

Structural determination of CatM- and BenM-EBD R156H variant regulators.

Since the R156H replacement affected regulation by CatM and BenM, the EBD structures of these variants were characterized. During construction of the plasmid to express BenM(R156H), site-directed mutation introduced the desired replacement and an accidental change, T157S. Since strains encoding BenM(R156H,T157S) regulated *benA*

expression similarly to BenM(R156H) (data not shown), the effect of the T157S replacement was inferred to be minimal. X-ray diffraction data sets were collected for two independent crystal forms of the BenM(R156H,T157S)-EBD variant without effectors. Their structures were deposited in the Protein Data Bank (PDB) with identification codes 2H99 (crystal form A) and 2H9B (crystal Form B). Both structures crystallized in space group $P2_12_12_1$ with similar cell constants and were solved by molecular replacement using as a model the coordinates of the previously solved BenM-EBD wild-type structure, PDB 2F7A (Ezezika *et al*., 2007a). As observed for the previous BenM-EBD structures, two subunits related by non-crystallographic two-fold symmetry were in the asymmetric units of both crystals, although the composition of ordered ions differed between the two structures. This difference in ion composition provided the impetus for determining the two otherwise isomorphous structures. In structure 2H99, an acetate molecule bound in the primary effector-binding site of one subunit (A), while this same position in its second subunit (B) was occupied by chloride and sulfate ions. In structure 2H9B, a sulfate ion was bound in the primary effector-binding site of one subunit (A), while the other subunit (B) bound a sulfate and a chloride ion. Both crystal forms diffracted to high resolution (approximately 1.85 Å). Details of the crystallographic data collection and refinement statistics are reported in Table 3.

In contrast to BenM(R156H)-EBD, CatM(R156H)-EBD was structurally characterized only in the presence of muconate. The variant structure (PDB 3GLB) was determined using the wild-type CatM-EBD structure (PDB 2F7B) as the molecular replacement model (Ezezika *et al*., 2007a). The CatM(R156H)-EBD structure had four subunits in the asymmetric unit (space group, $P2_12_12$) arranged as two independent

dimeric units. In contrast, the previously determined CatM-EBD structures all had single subunits in the asymmetric units that could create dimers by applying crystallographic two-fold operators. Muconate molecules were present in the primary effector binding sites of all the CatM(R156H)-EBD subunits. This variant crystal had a lower resolution limit of (2.8 Å).

Assessing structural differences

To compare the variant and wild-type EBD structures, individual subunit structures were aligned using invariant core analysis with the program Bio3D (Grant *et al.*, 2006). This analysis was consistent with the two domains of each EBD acting as rigid bodies that flex around a central hinge near residues 162 and 265. In Figure 3.2A, this hinge-like region corresponds to the orange and green beta-strands depicted behind muconate. To illustrate the structural effects of the R156H replacements in BenM and CatM, residues from only domain I were aligned using invariant core analysis. The full EBD structures were then overlaid onto the aligned domain I residues. In this fashion, domain II movement could be visualized relative to a fixed domain I, Figure 3.6.

To evaluate structural differences, the root mean square deviations (RMSD) of residues 90-161 and 267-302 of domain II were calculated between the unliganded BenM-EBD structure (2F6G, subunit A) and the other BenM-EBD structures using domain I-aligned proteins. A similar approach was used for CatM-EBD (residues 90-161 and 267-286). However, the calculated values were difficult to interpret and relatively small (with a maximum value of 1.76 Å), as reported in Tables A.1 and A.2 (Appendix A). The results were complicated by the different directions of relative movement

observed between domain I and domain II in different structures. Therefore, a more sophisticated method of evaluation was needed.

To improve the detection of functionally significant changes in the relative positions of the two domains of the EBDs, a domain motion analysis was performed with the program DynDom (Hayward, 1999, Hayward & Berendsen, 1998, Hayward & Lee, 2002). In this analysis, the movement of domain II with respect to a fixed domain I (containing the R156H replacement) was described by the displacement of the rigid body around a defined screw axis with a translation along this axis. Prior to the motion analysis, the individual BenM-EBD subunit structures were superimposed using core invariant residues within domain I. The CatM-EBD structures were similarly aligned and subjected to motion analysis. Domain movement can be evaluated by considering the orientation of the screw axis, as discussed later and shown in Figure 3.6B and 3.6D for BenM-EBD and CatM-EBD structures, respectively.

Discussion

BenM's response to benzoate requires key residues in the secondary effector-binding site

This investigation shows that benzoate-activated transcription and dual-effector synergism both depend on a BenM region distinct from the primary effector-binding site. The response to benzoate relies on two residues in a recently discovered secondary effector-binding site (Ezezika *et al*., 2007a). Individual or double replacements of R160 and Y293 demonstrated the functional significance of the secondary site. While these changes in BenM variants can abolish growth on benzoate and prevent benzoate-induced *benA* expression (Figure 3.3A), the response to muconate was retained and appears to

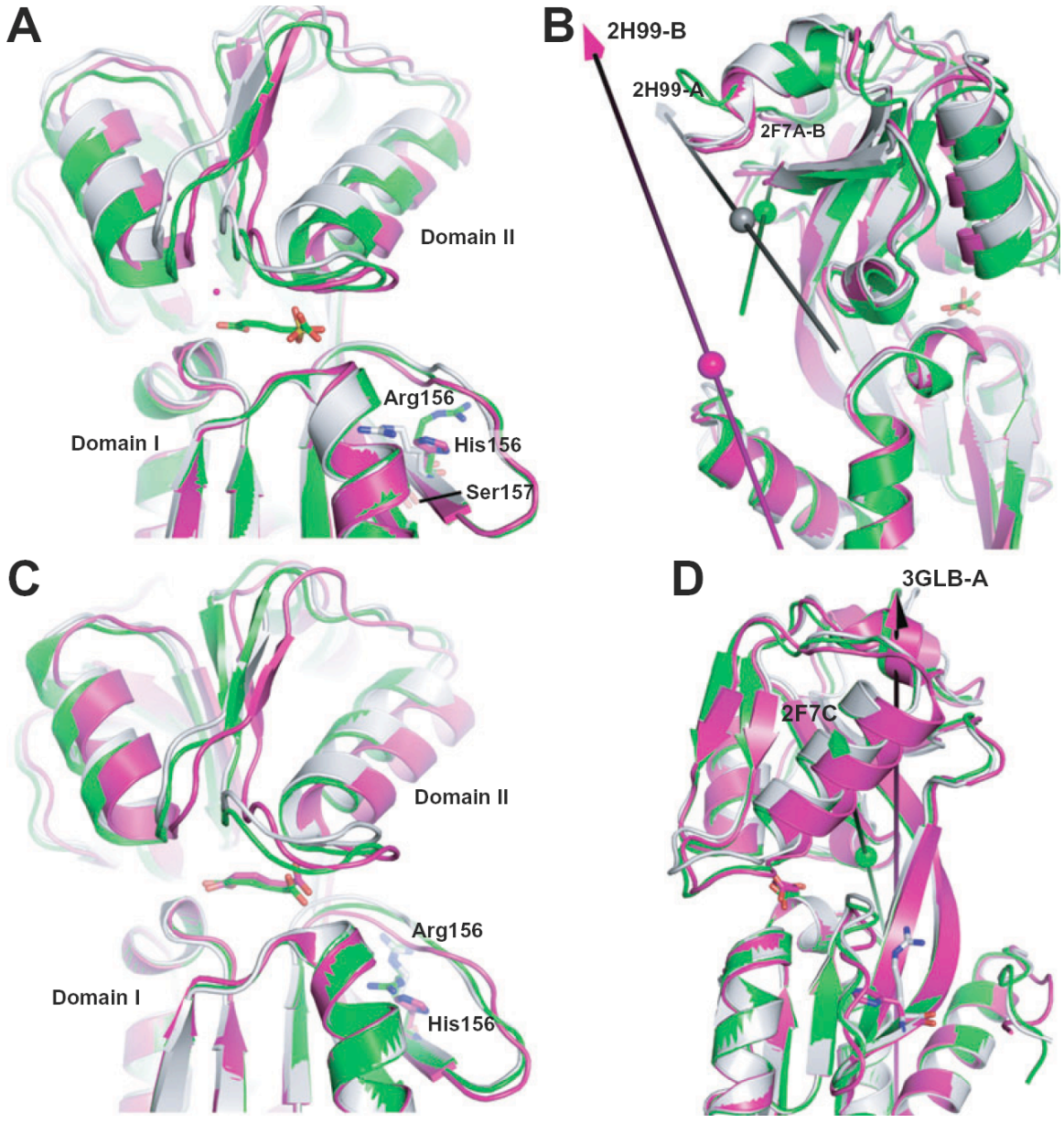

A
Domain II
Arg156
Domain I
His156
Ser157
B
2H99-B
2H99-A
2F7A-B
C
Domain II
Arg156
Domain I
His156
D
3GLB-A
2F7C

Figure 3.6: Structural effects of the R156H replacement in BenM and CatM. **A.** Using invariant core residues from domain I, three BenM-EBD structures were superimposed: unliganded wild-type (light grey, PDB ID 2F6G subunit A), muconate-bound wild-type (green PDB ID 2F7A subunit B), and the R156H,T157S variant without effectors (magenta, PDB ID 2H99 subunit B). Chloride (magenta sphere) and sulphate ions occupy the primary effector-binding site of the variant. **B.** The images from A are rotated 90° about the vertical axis. Arrows indicate the screw axes that describe the domain II motions of the BenM-EBD structures relative to the unliganded wild-type structure (2F6G, subunit A). Each axis is labeled with the ID of its structure and the subunit designation (A or B). Colors of the sphere and arrow point match that of the structure, except for variant 2H99-A, whose structure is not shown. **C.** A similar representation of CatM-EBD structures: unliganded wild-type (light grey, PDB ID 2F7B), muconate-bound wild-type (green, PDB ID 2F7C, subunit A) and muconate-bound R156H variant (magenta, PDB ID 3GLB, subunit A). **D.** A rotated view of C showing screw axes for CatM-EBD structures as described for B.

depend on the primary effector-binding site (Figure 3.3B). Moreover, the regulatory effects of the distinct primary and secondary effector-binding sites are interconnected (Figure 3.2B). Amino acid replacements designed to prevent benzoate from binding the secondary site caused benzoate to inhibit muconate-induced *benA* expression (Figure 3.3B). The absence of benzoate in the secondary site may improve the ability of benzoate to compete with muconate for binding to the primary site. Such competition is consistent with previous studies that show benzoate is able to bind in the primary site (Clark *et al.*, 2004b, Ezezika *et al*., 2007a).

Charge-dependent interactions of residues between the primary and secondary effector-binding sites form the basis of our model for synergism (Craven *et al.*, 2008, Ezezika *et al*., 2007a). According to this model, benzoate bound in the secondary site enhances the effect of muconate in the primary site due to interaction between the carboxyl group of benzoate and R160 (Figure 3.2). This interaction decreases the attraction between R160 and E162 and enhances the opportunity for interactions between E162 and R146. This latter interaction keeps R146 from shielding the negative charge of the adjacent carboxyl group of muconate, thereby strengthening the attraction between the negatively charged effector and the positive dipole moments of four nearby alpha helices. Therefore, in the presence of benzoate, muconate should be more effective in drawing together the two domains of the EBD into an activated conformation.

The R160M and R160K replacements might be expected to affect muconate-induced gene expression differently. Without benzoate in the secondary site, methionine could enhance interaction between E162 and R146 to augment muconate induction in a fashion that mimics the benzoate-R160 contact in wild-type BenM. In contrast, the

charge based attraction between K160 and E162 might resemble the effect of the R160 in the absence of benzoate, yielding lower-levels of muconate-induced transcription. While R160M was not observed to enhance gene expression to a greater extent than the R160K in cultures grown with high concentrations of extracellular muconate (Figure 3.3B), the only variant that allows growth on benzoate as the sole carbon source is BenM(R160M,Y293F). This result suggests that the R160M replacement is better able than R160K to activate muconate-dependent transcription in vivo.

Consistent with our regulatory model, the R160M replacement could enhance transcription via an increased affinity of the protein for muconate that is physiologically significant at low effector concentrations (Figure 3.4). The long lag time of ACN639 suggests that growth is limited by the time needed to generate muconate as an inducer (Table 2). It is not yet clear why the Y293F replacement is also required for growth on benzoate and for the early elevation of *benA* expression (Figure 3.4). Furthermore, it is not evident why the replacements at positions 160 and 293 reduced muconate-dependent *benA* expression relative to wild-type BenM (Figure 3.3B). Such observations emphasize the need to investigate structural differences and protein stability in addition to correlating ongoing functional and structural analyses.

Transcriptional activation without effectors: the importance of protein interfaces in the tetramer

Spontaneous mutants revealed the importance of residues 225 and 226. BenM(R225H) and BenM(E226K) activated high-level *benA* expression without exogenous inducers (Figure 3.5). In BenM-EBD structures, the side chains of R225 and

E226 are adjacent and lie at the center of the dimer interface on the two-fold symmetry axis (Figure 3.7C). This axis, present in all our structures, is crystallographic in some cases and non-crystallographic in others. Thus, the R225H and E226K replacements are likely to affect the dimer interface. Supporting this assumption, a comparably positioned residue in Cbl was shown by cross-linking studies to be required for oligomerization (Stec *et al*., 2006). Residues 225 and 226 are identified in a model of the tetrameric BenM-EBD structure (Figure 3.7). Although BenM and CatM did not crystallize as tetramers, the interface between neighboring subunits (the dimer interface) was common in all our EBD structures (Ezezika *et al*., 2007a, Ezezika *et al.*, 2007b). Furthermore, the organization of two BenM-EBD structures (2F8D and 2F97) in their crystal lattices mimicked a biologically active tetramer, as assessed by the remarkable similarity to tetrameric CbnR and DntR structures (Ezezika *et al*., 2007b, Muraoka *et al*., 2003, Smirnova *et al*., 2004). Mutations that cause constitutive activation in other LTTRs, such as OccR and OxyR, have also been predicted to affect the dimerization interface (Akakura & Winans, 2002a, Choi *et al*., 2001). Amino acid replacements may alter the relative orientations of the subunits or the local flexibility of the protein subunits and thereby affect the function of the oligomeric protein.

The R225H replacement may favor ionic interactions between neighboring side chains of H225 and E125 at the expense of those with E226 (Figure 3.7C). At this central location of the dimer interface, both the charge (pKa) and mass differences between

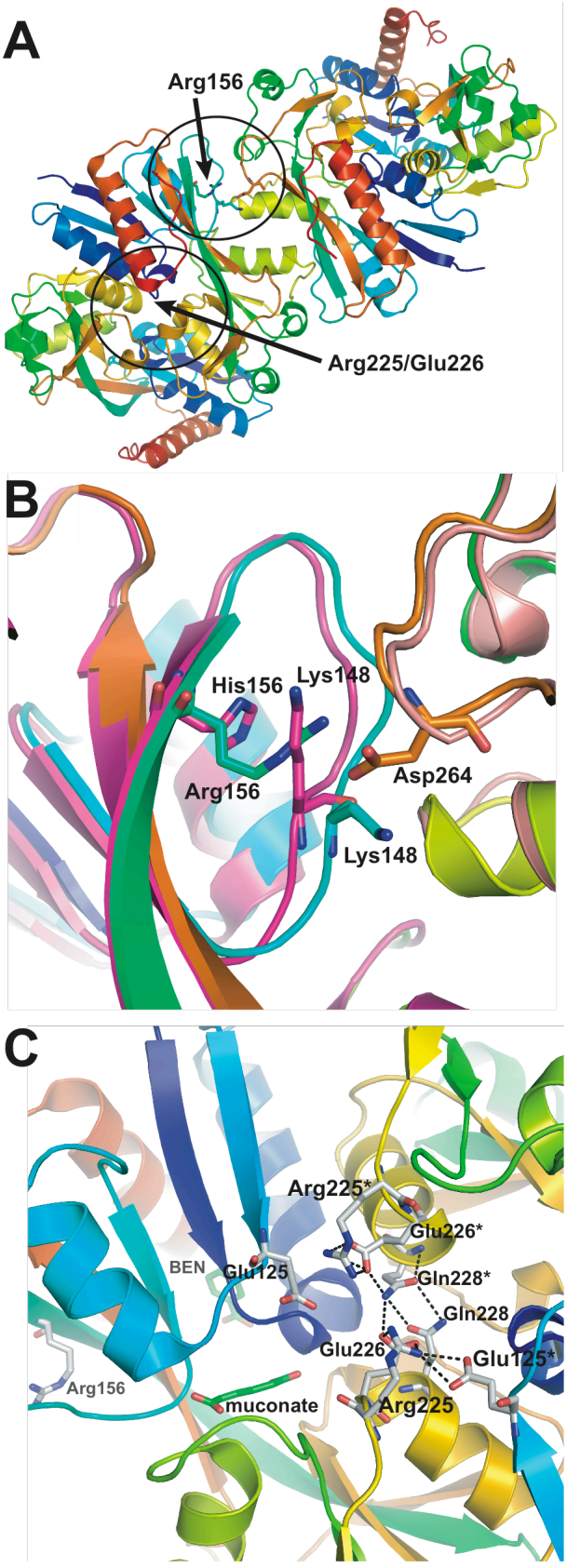

A
Arg156
Arg225/Glu226
B
His156
Lys148
Arg156
Asp264
Lys148
C
Arg225*
Glu226*
BEN
Glu125
Gln228*
Gln228
Glu226
Glu125*
Arg156
muconate
Arg225

Figure 3.7: Location of amino acids in BenM variants. A. The predicted tetrameric arrangement of subunits based on the expansion of crystallographic symmetry to a BenM-EBD structure (PDB ID 2F97) (Ezezika *et al*. 2007b). Each subunit is colored blue (N-terminus) to red (C-terminus). Circles highlight the location of variant residues. **B.** An enlargement of the top circled area in A depicting the region near residue 156. The BenM(R156H,T157S)-EBD structure (PDB ID 2H99) is superimposed on that of the wild-type (PDB ID 2F97) using resides within 6Å of the interface. **C.** An enlargement of the bottom circled area in A depicting the region near residues 225 and 226. Benzoate (BEN) in the secondary effector-binding site and muconate in the primary site are shown in green. Arg156 is shown for reference. Hydrogen bonds between ionic residues are depicted as dashed lines. Symmetrically related residues from the twofold-related subunit are marked with an asterisk.

arginine and histidine are likely to affect the association properties between adjacent proteins. With the E226K replacement, a hydrogen bond between K226 and E125 might pull the domains together and mimic effector-mediated activation. Without effectors, the BenM(E226K) variant activates maximal levels of transcription that are not significantly enhanced by benzoate or muconate (Figure 3.5). This regulation enables ACN866 to grow rapidly on benzoate with a short lag time (Table 2).

BenM(R156H) also activates transcription without effectors and allows rapid growth on benzoate (Figure 3.5, and Table 2). The importance of R156 was first found in CatM (Neidle *et al*., 1989, Ezezika *et al*., 2006). The R156H replacement, which lies at the interface between dimers (the tetramer interface), may weaken the interaction with D264 on the adjacent subunit (Figure 3.7A and B). Furthermore, the R156H replacement may alter the loop of residues 148 to 156 and thereby impact the relationships between subunits of the oligomeric protein. The structure of the variant showed substantial repositioning of lysine at position 148 (Figure 3.7B). When H156 is present, lysine moves into the adjacent pocket normally occupied by the side chain of arginine. Because the BenM(H156H,T127S)-EBD structures did not pack using the tetrameric interface, we cannot identify the effects on oligomer reorientation. However, since this tetrameric interface was not used for packing in the crystals, alterations in tetramer formation could not have caused the changes observed in the relative positions of the EBD domains in the wild type and the R156H variants (described next). There were fundamental changes in each subunit.

Protein conformations associated with LTTR-activated transcription and ligand binding

In LTTRs, effectors are predicted to bring the EBD domains together by binding between them (Ezezika *et al*., 2007a). Muconate in the primary site of BenM-EBD (green) resulted in a downward tilting of the right side of domain II towards domain I compared to the structure without ligands (gray, Figure 3.6A). Similarly, both domains of the variant (magenta) are drawn together despite the absence of muconate (Figure 3.6A). Even without the context of the functional tetramer, the variant EBD appears to assume a conformation similar to the inducer-bound wild type. Since the variant contains chloride and sulfate ions in the primary effector-binding site, BenM(R156H) may activate transcription nonspecifically in response to bound anions.

Structural comparisons were complex, and domain movement was not limited to rotation about a hinge. Motion analysis was used to determine the position of a screw axis to describe simultaneous rotational and translational movements of domain II given a fixed position of domain I, using the regulator without ligands as a reference (Figure 3.6B). The movement of domain II towards domain I can be visualized using a right-hand rule with the thumb pointing in the direction of the arrow (translation) such that the fingers curl around in the direction of rotation.

The position of an axis representing simple rotation about an inter-domain hinge is best matched by the wild-type structure with muconate and benzoate bound in the primary and secondary effector sites, respectively (2F7A-B, Figure 3.6B). In this inducer-bound structure, domain II rotates 5.5° with minimal translation (0.21 Å) about an axis that passes through domain II near residues 161 and 267 (Figure 3.6B). The movement parameters for all our structures are given in Tables A.1 and A.2 (Appendix A). The

BenM(R156H,T157S)-EBD variants deviated from a perfect hinge motion to a greater extent than the inducer-bound wild-type regulator. For each variant subunit, the position of its screw axis depended on the particular ions in its primary effector-binding site (Table 4). For the structures with both chloride and sulfate ions in this site, 2H99-B (Figure 3.6B) and 2H9B-B (Figure A.1), the rigid-body motion of domain II is skewed towards the location of the R156H replacement. While the significance of this twist is unknown, the R156H variants were more akin to the wild-type structure bound to muconate than to those that lacked this inducer. This conclusion is supported by assessment of the screw axis vectors and other displacement values (such as the distances between the screw axis and the line joining the centers of mass). Thus, the rigid-body motion of the domains in the BenM(R156H,T157S)-EBD variant imitated the conformational changes that typically occur upon muconate binding.

The movement analysis for CatM-EBD was more complicated. First, the wild-type structure (2F7B) contained sulfate and chloride ions in the primary effector-binding site that appeared to draw the EBD domains together in a partially closed form intermediate between the BenM-EBD structures with and without inducers (2F7A-B and 2F6G-A, respectively). Second, we were only able to obtain data for muconate-bound CatM(R156H)-EBD. Nevertheless, the comparisons were informative. The EBD domains of the variant structure (magenta) were closer together when clamping down on muconate than even the inducer-bound wild type (green, Figure 3.6C). Since muconate-dependent induction by this R156H variant is higher than for wild-type CatM (Neidle *et al*., 1989), our data support the model that a transcriptionally active regulator has a conformation in which the two EBD domains are drawn together. Although the conformational change

within each protein subunit may appear to be subtle, the propagation of small changes can result in substantial changes within the tetramer (Ezezika *et al*., 2007a).

The screw axes of the CatM(R156H)-EBD structures (3GLB, subunits A, C, and D) show that relative to the wild-type structures, domain II is rotated more prevalently toward the R156H replacement (Figure 3.6). Subunit B of the CatM(R156H)-EBD structure displays the most exaggerated rotation toward the replacement. A sulfate ion, unique to this subunit, may bridge His160 and His297 and distort the conformation. Overall, in the structures of both BenM(R156H,T157S) and CatM (R156H), the rigid-body domain motions resemble those of the muconate-bound wild-type proteins, but with a slight turn to bring domain II toward residue 156. Such a change most likely affects the relative orientations of these domains in the formation of the functional oligomeric proteins.

Dual-effector synergism and relevance to the widespread LTTR family

Although BenM is the only regulator in which distinct effector-binding sites and transcriptional synergism are known, there are likely to be related examples of multi-effector transcriptional control. A database search revealed approximately 60 BenM-like homologs in which key residues in the secondary effector-binding pocket were conserved (data not shown). Completely conserved residues included R146, R160, and E162, which may mediate transcriptional synergism, as well as R160 and Y293, which may interact with a ligand. These putative regulators also contained conserved residues in a pocket resembling the muconate-binding primary effector-binding site of BenM and CatM. These homologs were identified in diverse bacteria including different species of

Burholderia, Ralstonia, Psychrobacter, Polaromonas, Methylobacterium, Sphingomonas, Bordatella, Xanthobacter, Paracoccus, and *Acinetobacter*.

The study of a BenM homolog from *Acinetobacter calcoaceticus* PHEA2, 84% identical to that of ADP1, did not identify synergism (Zhan *et al.*, 2008). However, this investigation failed to assess metabolism and inducer uptake properly, raising questions about the conclusions. The effector responses of many LTTRs involved in aromatic compound degradation have proven difficult to study. Nevertheless, the potential use of such regulators for purposes such as biosensing and bioengineering are driving efforts to create regulators with increased sensitivity to specific inducers and/or more varied effector-binding profiles (Cebolla *et al.*, 1997, Cebolla *et al.*, 2001, Lonneborg *et al.*, 2007, Smirnova *et al*., 2004). Our structure-function studies of BenM and CatM may be helpful in extending such biotechnology applications for the LTTR family.

Experimental procedures

Bacterial strains and growth conditions

A. baylyi strains (Table 3.1) and *Escherichia coli* were grown in Luria-Bertani (LB) broth at 37 °C with shaking (300 rpm) (Sambrook *et al.*, 1989). Alternatively, *A. baylyi* was grown in minimal medium with succinate (10 mM), anthranilate (2mM), benzoate (2 mM), or muconate (2 mM) as the carbon source (Shanley *et al.*, 1986). *E. coli* DH5α (Invitrogen) and XL1-Blue (Stratagene) were used as plasmid hosts. Antibiotics were added, as needed, to the following concentrations: ampicillin, 150 μg/ml; kanamycin, 25 μg/ml; streptomycin, 13 μg/ml; spectinomycin, 13 μg/ml. For *A. baylyi* growth curves, 5 ml cultures were inoculated with succinate-grown colonies. After

overnight growth on benzoate, 1 ml was used to inoculate 100 ml of benzoate medium. Growth was assessed by turbidity and measured spectrophotometrically ($OD_{600\ nm}$).

Site-specificic mutagenesis of benM and generation of A. baylyi *strains via allelic exchange*

Site-directed mutagenesis of plasmid-borne *benM* (QuikChange, Stratagene) was used to encode desired amino acids. Template plasmids, pBAC7 or pBAC433, were used in PCR reactions with mutagenic primers (Table A.3, Appendix A). *Dpn*I-treated PCR products were transformed into XL1-Blue cells, and plasmid-containing colonies were selected. DNA sequencing confirmed nucleotide substitutions on plasmids. Plasmid-borne alleles were used to replace chromosomal genes (Collier *et al*., 1998). To aid the introduction of altered *benM* DNA, a counter-selectable marker disrupted chromosomal *benM*. Plasmid pBAC709 was made by ligating the ~3.65 kbp *sacB*::Km^r cassette from PRMJ1 (Jones & Williams, 2003) into the *Sal*I site of *benM* on pBAC11. pBAC709 was linearized with *Alw*N1 and used to transform ADP1, generating ACN637. To replace the chromosomal *sacB* marker with modified *benM* alleles, appropriate plasmids were linearized with *Xho*I or *Pvu*II and used to transform ACN637. Desired transformants were selected by growth at 30 °C in the presence of 5% sucrose. In resulting strains, the chromosomal *benM* was PCR amplified and confirmed by DNA sequencing.

β-galactosidase assays to measure benA::*lacZ expression*

A *benA*::*lacZ* transcriptional fusion was introduced into the chromosome of *A. baylyi* strains by allelic exchange with DNA from pBAC54 linearized with *Xmn*I. Unless

otherwise noted, strains were grown overnight in LB with kanamycin. The following day, strains were subcultured into LB with no inducer or the following: benzoate (1 mM), muconate (1 mM), or both (0.5 mM each). Cell density was measured at 600 nm, and LacZ activity was assayed when cultures entered stationary phase (Miller, 1972). Using the FlourAce β-galactosidase reporter kit (Bio-Rad), the hydrolysis of 4-methylumbelliferyl-galactopyranoside to the product 4-methylumbelliferone was detected with a TD-360 minifluorometer. For assays during growth on anthranilate, cultures were first grown overnight in 5-ml LB with kanamycin. The following morning, 100 µl of culture were diluted into 5 ml of minimal medium with 1mM anthranilate as the sole carbon source. LacZ activity was measured at specific times following inoculation.

Transformation assay used to confirm the ben+ growth phenotype of ACN639

To assess the ben+ phenotype of ACN639, transformation assays were performed. ACN824, encoding the Y293F replacement, was used as the recipient with donor DNA from plasmids. pBAC711, pBAC771, and pBAC772 were linearized by digestion with *Xho*I, and pBAC780 with *EcoR*I, before being used to transform ACN824 (Table 3.1). The same transformations were done using ACN812, encoding BenM(R160M), as the recipient strain. Transformants grew on benzoate only when an allele could be generated to encode BenM(R160M,Y293F).

Expression and purification of BenM- and CatM-EBD variants

The effector-binding domains of BenM(R156H,T157S) and CatM(R156H) were purified from *E. coli* BL21(DE3) Gold cells (Stratagene) harboring expression plasmids

pBAC698 or pBAC683, respectively. For expression, *E. coli* cultures were grown at 37°C to an OD_{600} of ~0.2 at which time isopropyl-β-D-thiogalactopyranoside (0.2 mM) was added for induction. Cells were harvested after overnight growth and suspended in lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10% [v/v] glycerol, 5 mM imidazole, pH 7.9) at 4 °C. Cells were lysed by French Press (15,000 psi) at 4 °C. The cell lysate was centrifuged at 15,000**g** for 15 min at 4 °C and the clarified cell extract was applied to a Hi-Trap 5 ml metal chelating column (GE Biosciences) charged with nickel. Purified BenM(R156H,T157S)-EBD samples were dialyzed twice against 20 mM Tris-HCl (pH 7.9), 500 mM NaCl, 10% (v/v) glycerol and concentrated to ~14 mg/ml. Purified CatM(R156H)-EBD was dialyzed against 20 mM Tris-HCl, 250 mM imidazole (pH 7.9), 500 mM NaCl, 10% (v/v) glycerol. Imidazole and glycerol were added to increase protein solubility.

Crystallization and X-ray analyses

Initial high-throughput crystallization screens at the Hauptman-Woodward Institute used microbatch under oil methods at 298K (Luft *et al.*, 2001). Screens were done with and without benzoate and/or muconate (100 mM) in the protein solution. Conditions yielding crystals were optimized in-house with the microbatch under oil method at 15 °C or 25 °C with 2 µl protein and 2 µl precipitant. Two conditions produced BenM(R156H,T157S)-EBD crystals. Crystal Form A came from a precipitant solution of 0.015 M magnesium acetate, 0.05 M sodium cacodylate, 1.7 M ammonium sulfate, pH 6.0; and Crystal Form B was derived from 2.0 M ammonium sulfate as precipitant. CatM(R156H) crystals grew from a precipitant of 1.6 M ammonium sulfate, 0.1 M citric

acid, pH 4.0 with 100 mM muconate (from a 500 mM, pH 7.0 stock solution). Protein samples were centrifuged for 5 min at 16000 *g* and allowed to equilibrate to room temperature before crystallization trials. Crystals for X-ray diffraction studies grew within one to two weeks.

Diffraction data for BenM(R156H,T157S)- and CatM(R156H)-EBD were collected at the Southeast Regional Collaborative Access Team (SER-CAT) at the 22-ID and 22-BM beamlines at the Advanced Photon Source, Argonne, Il with 0.5° oscillations and wavelength of 1.0 Å. Data were processed with beamline versions of HKL2000 (Otwinowski & Minor, 1997). Structures were determined as before (Ezezika *et al*., 2007a, Ezezika *et al*., 2007b) by molecular replacement using MOLREP in the CCP4 suite (1994) with known EBD coordinates (PDB accession 2F6G and 2F7A for the BenM variant and 2F7B and 2F7C for the CatM variant) as search models. COOT (Emsley & Cowtan, 2004) was used to adjust local differences in the structures, and atomic refinement performed with REFMAC (Murshudov *et al.*, 1997). Refinement of the four non-crystallographically-related subunits of CatM(R156H) used medium NCS restraints on four groups defined by residues 90-155, 157-219, 221-275, and 277-296. Residues 297-307 at the C-termini were ordered to different degrees in each subunit and assumed several conformations.

Invariant core residues were identified and aligned in the program Bio3d (Grant *et al*., 2006) with a 0.5 $Å^3$ core cutoff. Atomic coordinates were aligned with different combinations of structures and subsets: full EBD subunits, domain I alone, domain II alone (residues), and as separate homologs or together. The overall root mean square (rms) deviations of the domain I residues were small (~0.6 to 1.2 Å) and not substantially

increased in the local regions of the replacements in variant structures. For graphical representations, BenM-EBDs were aligned on core residues from domain I (91-98, 100-109, 111-112, 121-127, 130-146, 154-60,268-274 and 281-293). CatM-EBDs were aligned on core residues (91-123, 125, 127-146, 154-154, 156-161, 268-274, 277-277, 279-293, and 295-295). The molecular graphics program PYMOL (DeLano, 2002) was used for visualization, figure preparation and rms calculations of aligned molecules.

Domain motions were evaluated using the Domain Select option of web-based Protein Domain Motion Analysis program DynDom (Hayward & Berendsen, 1998, Hayward & Lee, 2002). For domain analysis, the pre-aligned subunits from invariant core analysis were truncated at the N- (residue 90) and C-termini (residue 302 for BenM, residue 296 for CatM), alternative side chain conformations were removed, and the subunits were entered with domain I as the fixed domain (for BenM, residues residues 90-161 and 267-302; for CatM, residues 90-161 and 267-286) and domain II (residues 162-266) as the moving domain. Screw axes were visualized as CGI objects by reformatting the Rasmol (Sayle & Milner-White, 1995) output of the DynDom program for visualization in the program PYMOL (DeLano, 2002). Screw axis vectors and center points were similarly calculated using a python script with data from the Rasmol output.

CHAPTER 4

DOUBLE TROUBLE: MEDICAL IMPLICATIONS OF GENETIC DUPLICATION AND AMPLIFICATION IN BACTERIA[3]

[3] Craven, S.H. and E.L. Neidle (2007) *Future Microbiology* **2**(3) 309-321. Reprinted here with permission of publisher.

Summary and Introduction

Summary

Gene amplification allows organisms to adapt to changing environmental conditions. This type of increased gene dosage confers selectable benefits, typically by augmenting protein production. Gene amplification is a reversible process that does not require permanent genetic change. Although transient, altered gene dosage has significant medical impact. Recent examples of amplification in bacteria, described here, affect human disease by modifying antibiotic resistance, the virulence of pathogens, vaccine efficacy, and antibiotic biosynthesis. Amplification is usually a two-step process whereby genetic duplication (step one) promotes further increases in copy number (step two). Both steps have important evolutionary significance for the emergence of innovative gene functions. Recent genome sequence analyses illustrate how genome plasticity can affect the evolution and immunogenic properties of bacterial pathogens.

Introduction

Gene amplification is a common process that has a large impact on human health and disease. It is an important factor underlying the development of some types of cancer as well as resistance to many chemotherapeutic agents (Albertson, 2006b, Yasui *et al.*, 2004). Amplification causes elevated levels of the multidrug resistance-1 transporter in *Plasmodium falciparum* leading to clinically significant drug failure in the treatment of malaria (Valderramos & Fidock, 2006). Here we focus on examples of gene duplication and amplification in bacteria that significantly affect diverse aspects of pathogenesis, virulence, antibiotic resistance, and vaccine efficacy. Detailed analyses of bacterial DNA

rearrangements continue to improve our understanding of the molecular mechanisms that mediate and promote genome plasticity. Moreover, the influence of such genetic changes on the clinical outcome of infectious diseases is a burgeoning area of investigation.

In this review, we focus on recently published examples of medically relevant gene amplification in bacteria and update previous discussions (Reams & Neidle, 2004b, Romero & Palacios, 1997). Related issues are also addressed, such as the role of gene duplication in the evolution of bacterial pathogens, the role of repetitive DNA in genome organization, and the role of small tandem nucleotide repeats in antigenic variation and methods of molecular epidemiology. Mobile genetic elements involved in horizontal gene transfer enhance the propensity for gene amplification by providing different copies of identical DNA sequences as targets for genetic recombination. This interconnection between horizontal gene transfer and gene amplification influences phenomena ranging from multiple drug resistance to the evolution of pathogenicity islands (Reams & Neidle, 2004b). Renewed awareness of the medical significance of genetic duplication highlights the importance of characterizing and understanding its molecular basis.

Gene Amplification via Recombination Between Direct Repeats

Gene amplification refers to an increase in the number of copies of a DNA segment relative to neighboring regions. Homologous recombination between identical sequences (direct repeats) can initiate the process (Figure 4.1). Sister chromosomal recombination can generate tandem duplication, and additional rounds of homologous recombination can further increase the copy number of the segment, termed the amplicon (Figure 4.1).

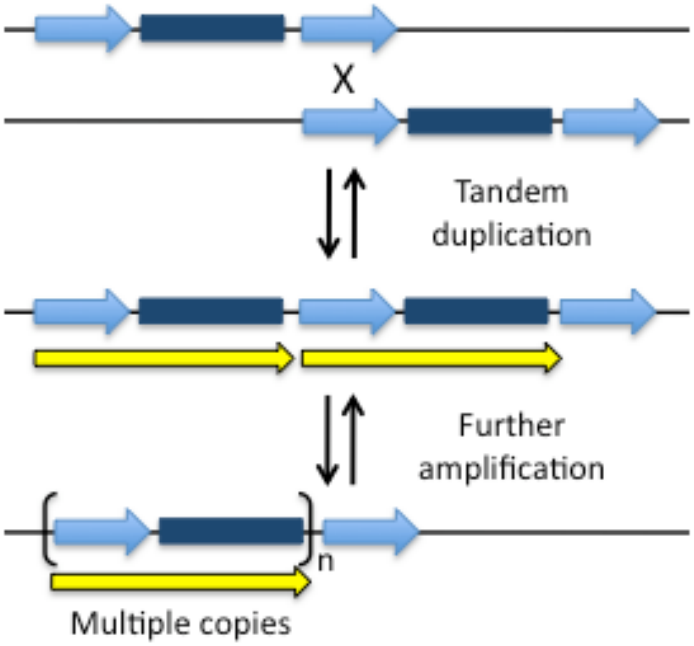


Figure 4.1: One method of reversible gene amplification. Homologous recombination (X) can occur between different copies of a directly repeated DNA sequence (blue arrow). This process of tandem duplication creates a larger, directly repeated sequence (yellow arrow) that can be involved in additional rounds of homologous recombination to generate a tandem array with multiple copies of the entire segment, which is termed the amplicon.

However, an alternative result of homologous recombination between direct repeats is deletion of the intervening DNA. Thus, amplification is reversible, dynamic, and frequently transient. In the absence of selective conditions favoring its retention, amplified DNA usually reverts to a haploid state.

Various types of repeated genomic sequences affect duplication frequencies. For example, identical sequences in different ribosomal RNA (*rrn*) operons contribute to the surprisingly high duplication frequency of some regions of the *Escherichia coli* chromosome (Anderson & Roth, 1981). Consistent with the ability of specific sequences to influence recombination events, there is extensive variation in the duplication frequency of different *E. coli* genes depending on genomic position, ranging from approximately 10^{-6} to 10^{-2} per cell (Anderson & Roth, 1981). Moreover, direct repeats are frequently provided by insertion sequences (IS), which are transposable genetic elements commonly found in prokaryotic genomes (Siguier *et al.*, 2006). The presence of IS copies in different locations and their relatively large size (approx. 800 to >3,000 bp) make these elements ideal targets for homologous recombination. In some cases, transposition precedes recombination between different IS copies (Reams & Neidle, 2003). Genes affecting pathogenicity are frequently carried by mobile elements that undergo this type of recombination, as discussed previously (Reams & Neidle, 2004b).

Examples of IS-mediated Amplification Affecting Virulence, Vaccine Failure, and Multiple Drug Resistance

A capsule-encoding gene cluster of *H. influenzae* is mobilized as a compound transposon carrying the *cap* genes between direct repeats of IS*1016.* These repeats undergo homologous recombination and further amplification that affect the amount of capsule, an important virulence determinant (Kroll *et al.*, 1991). Most invasive isolates of *H. influenzae* serotype b (Hib) possess a large, approximately 18 kb, *cap*-gene chromosomal duplication. A 1.2 kb deletion at one end of the duplication exerts selective pressure to maintain two *cap*-gene copies, which enhances capsule production (Kroll *et al*., 1991). Further amplification of this locus generates hyper-encapsulated variants with decreased susceptibility to complement-mediated opsonization and lysis (Cerquetti *et al.*, 2005). In one strain harboring six copies, the *cap* b locus was implicated in a severe clinical presentation leading to fulminant septicemia and patient death (Cerquetti *et al.*, 2006a).

Although routine use of Hib conjugate vaccines dramatically decreased the occurrence of invasive Hib disease in children worldwide, a recent increase has been noted in some immunized children (Cerquetti *et al*., 2005, Cerquetti *et al.*, 2006b). Children with vaccine failure are more likely to be infected with a multiple-copy strain (≥ 3) than are unvaccinated children (Cerquetti *et al*., 2005, Cerquetti *et al*., 2006b). In an alarming trend, vaccination appears to exert positive selection for virulent strains harboring amplified *cap* b loci (Cerquetti *et al*., 2006b). The amplified state is unstable in

laboratory culture, yet amplified strains are selected in humans because they withstand host defenses such as serum factors and antibodies to capsular polysaccharide.

Another alarming situation involves invasive disease caused by isolates of *H. influenzae* serotype a (Hia) with stable *cap* duplication (Kapogiannis *et al.*, 2005). Typically, non-b capsular serotypes of *H. influenzae* do not cause disease in children. Recently, two Hia isolates that caused severe pediatric disease were found to carry a deletion at one end of a duplicated *cap* locus, resembling that of Hib invasive strains (described above) (Kapogiannis *et al*., 2005). The possibility is raised that stable duplication and further amplification of capsule loci increase the virulence of non-b *H. influenzae* strains. Such strains could fill the niche left vacant by the vaccine-reduced nasopharyngeal carriage of Hib. This kind of serotype replacement could render the Hib vaccine less effective in preventing disease. Careful strain typing and monitoring of the genotype and copy number of *cap* loci remain important.

Drug resistance and multiple copies of the acrAB *genes of* Escherichia coli

Another example of IS-mediated gene amplification confers multiple antibiotic resistance (Mar) in *E. coli* (Nicoloff *et al.*, 2007, Nicoloff *et al.*, 2006). In recent studies, spontaneous mutants were selected in the presence of 4 μg/ml of tetracycline. Of thirteen mutants examined, all displayed a Mar phenotype characterized by low-level resistance to compounds from different antibiotic families. Such resistance depends on an active AcrAB-TolC multidrug efflux pump able to expel a variety of chemically unrelated substrates. In several independent mutants, amplification of the *acrAB* genes caused the

Mar phenotype. Duplication was initiated by homologous recombination between IS elements on either side of *acrAB*.

Two copies of IS*5* create direct repeats on either side of a 300 kb region carrying *acrAB*. In one isolate, recombination between these elements initiated duplication and further *acrAB* amplification. Similarly, recombination between two IS*3* elements in another mutant generated a 252 kb amplicon carrying the *acrAB* genes. Multiple amplicon copies persist under selective conditions, with an estimated two to three copies per chromosome after overnight growth with nalidixic acid. In contrast, when selective pressure is removed, amplification is lost along with the *mar* phenotype (Nicoloff *et al*., 2007, Nicoloff *et al*., 2006).

In another amplification mutant, duplication is preceded by transposition of IS*186* into the protease-encoding *lon* gene, a hotspot for this IS element (Nicoloff *et al*., 2006). Next, recombination with another chromosomal IS*186* element generates a 149 kb *acrAB* amplicon. Interestingly, most of the spontaneous Mar mutants also carry *lon*::IS*186* mutations. While *lon* mutation alone does not produce significant multi-drug resistance, it does enable slow growth and a fitness advantage relative to the wild-type strain under the selection conditions used. Furthermore, *lon* mutations increase IS transposition and genome rearrangements (Nicoloff *et al*., 2007).

The clinical relevance of *lon* mutation is not clear. Nevertheless, amplification of the *acrAB* region appears to be an important mechanism of *E. coli* drug resistance that does not depend on *lon* inactivation (Nicoloff *et al*., 2007, Nicoloff *et al*., 2006). Amplification of this region in combination with another mutation in *marR*, encoding a *mar*-regulon repressor, generates clinically relevant levels of drug resistance. Cells with

IS*3*- and IS*5*-mediated duplications are present in unselected populations (Nicoloff *et al*., 2007). These cells with two copies of the *acrAB* region provide a subpopulation from which high level amplification mutants could readily arise under clinical conditions.

Examples of Gene Duplication via Illegitimate Recombination

Gene amplification in prokaryotes is typically characterized as a step-wise process initiated by tandem duplication and followed by subsequent amplification. In contrast to the examples discussed above, the initial duplication step may occur without homologous recombination between long stretches of direct repeat sequences. In a seminal paper demonstrating the connection between drug resistance and gene amplification, recombination occurred between segments of identical sequence that were only 12 bp in length (Edlund & Normark, 1981).There are numerous cases where duplication results from RecA-independent recombination between sequences with short or no regions of sequence identity, i.e. illegitimate recombination (Francia & Clewell, 2002, Reams & Neidle, 2004a, Seoane *et al.*, 2003). Several mechanisms may be involved in the amplification step, including a rolling-circle process that enables rapid increases in amplicon copy number (Yanai *et al.*, 2006). There is further discussion about mechanisms underlying amplification in review articles and references therein (Reams & Neidle, 2004b).

Examples of illegitimate recombination and further amplification that affect drug resistance and enhanced antibiotic production

Yersinia enterolitica strains harbor β-lactamases that confer intrinsic resistance to β-lactam drugs such as ampicillin. In the presence of high concentrations of this antibiotic, growth of *Y. enterolitica* Y56 yields mutants with ampicillin-resistance levels as high as twenty times greater than in the wild type (Seoane *et al*., 2003). Characterization of these hyper-resistant mutants revealed tandem amplification of an approximately 28 kb chromosomal DNA fragment containing the β-lactamase-encoding *blaA* gene. High-level drug resistance correlates with the presence of at least five copies of the 28 kb amplicon. Consistent with amplification-based resistance, the hyper-resistant phenotype is unstable. The basal level of ampicillin resistance returns after the antibiotic is withdrawn. The DNA segments involved in the initial tandem duplication do not have significant stretches of sequence identity. However, these DNA segments have weak similarity to DNA gyrase recognition sites in *E. coli*. This observation raises the possibility that *blaA*-duplication in *Y. enterolitica* could involve DNA gyrase-mediated illegitimate recombination (Seoane *et al*., 2003). In *E. coli*, the exchange of DNA gyrase subunits in a two-enzyme complex has been proposed as a model to explain how this enzyme catalyzes the breakage and rejoining of different DNA double stranded regions in an illegitimate recombination reaction (Ikeda *et al.*, 2004).

Another example of amplification-based drug resistance occurs on pAMα1, a 9.75 kb multicopy plasmid of *Enterococcus faecalis* (Francia & Clewell, 2002). In this case, tandem amplification of a 4.1 kb amplicon carrying the *tetL* gene enhances resistance to tetracycline. Selective conditions lead to the maintenance of a tandem array

of as many as eight amplicon copies on a single plasmid molecule. The first step in amplification, tandem duplication, is mediated by recombination between sequences flanking *tetL*. These flanking sequences are relatively short and have low levels of sequence identity. Consistent with the inference that these sequences do not undergo homologous recombination, duplications form in a mutant lacking RecA. However, further amplification depends on RecA (Francia & Clewell, 2002). Interestingly, tandem duplication appears to involve site-specific recombination mediated by plasmid-encoded relaxases. These enzymes catalyze both DNA cleavage and ligation reactions in conjugative plasmid mobilization, so it is not surprising that they are capable of site-specific recombination. Medical concerns are raised by this type of amplification since clinical isolates of enterococci commonly acquire tetracycline resistance from plasmids (Clewell, 1990).

In a related example, amplification not only increases resistance, but also leads to enhanced antibiotic production in the streptomycete, *Streptomyces kanamyceticus*. Streptomycetes are frequently used for the industrial production of antibiotics such as kanamycin. To develop strains for this purpose, screening for antibiotic-overproducing mutants is often conducted after selection for resistance to the same antibiotic. This selection ensures that the bacterium can withstand the toxic effects of large amounts of its own secondary metabolite. Using this approach in *Streptomyces kanamyceticus*, amplification was discovered of the entire chromosomal kanamycin biosynthetic gene cluster (Yanai *et al*., 2006). In this case, several genes for antibiotic resistance and self-protection are located within a large cluster that includes biosynthetic genes. Tandem amplification of the cluster increases resistance to kanamycin and augments its

bioproduction in a fashion that depends on the copy number of a 145 kb amplicon (Yanai *et al*., 2006). DNA from the strain contains a mixture of genomes each with varying numbers of amplicons in the tandem array. Detection of a large amplified region >5 Mbp suggests that amplification can exceed 36 copies.

In this example, site-specific recombination generates the initial tandem duplication (Yanai *et al*., 2006). Recombination occurs between two distal copies of a six-nucleotide sequence. Additionally, a different rearrangement involving a deletion of DNA in the same region was detected. DNA analysis also identified repetitive sequences in this region that belong to a family of clustered regularly interspaced short palindromic repeats. The roles of these repeats in rearrangements of the kanamycin gene cluster remain unclear. However, further amplification of the duplication may arise via homologous recombination and/or a rolling circle replication method (Yanai *et al*., 2006). Additional studies to clarify the amplification mechanism(s) in *S. kanamyceticus* may provide new insight into methods for increasing the productivity of industrial strains for antibiotic synthesis.

Interplay between gene amplification and additional mutations

In many of the examples described above, amplification is not the sole genetic alteration. In some situations, an initial mutation increases the probability that an amplification mutant will confer a selectable phenotype. Such is the case for *lon* or *marR* mutations with *acrAB* amplification (Nicoloff *et al*., 2007, Nicoloff *et al*., 2006). Moreover, amplified DNA provides a large genetic target for alteration, multiple sites for recombination, and functional redundancy. Therefore, amplification increases the

opportunity for new mutations to accrue within a multi-copy region. In the case of *cap* amplification in *H. influenzae*, a deletion within the duplicated region is retained because it stabilizes a beneficial configuration (Kroll *et al*., 1991). In streptomycetes, the frequent accompaniment of amplifications by large deletions may reflect limitations in the amount of DNA that can be packaged in a spore (Yanai *et al*., 2006). As described below, the interplay between gene amplification and other types of mutation profoundly affects the selection and evolution of new traits.

The fitness costs of amplification and drug resistance: can antibiotic treatments be optimized to prolong the effective control of bacterial diseases?

Amplification can enhance drug resistance by directly increasing the production of proteins that mediate resistance (Francia & Clewell, 2002, Nicoloff *et al*., 2007, Seoane *et al*., 2003). In a variation on this theme, gene amplification indirectly facilitates resistance to actinonin by improving an alternative pathway that bypasses the drug-blocked reaction (Nilsson *et al.*, 2006). This reaction removes the formyl group from the initiating methionine in nascent peptides (Figure 4.2). By preventing deformylation, actinonin inhibits growth and causes toxic levels of formylated polypeptides to accumulate. However, actinonin-mediated toxicity can be alleviated by mutations that reduce formylation of the initiator methionyl-tRNA (tRNAi), such as mutations in *folD* or *fmt* (Figure 4.2).

Without formylation, protein synthesis must initiate with an unformylated tRNAi. As a result, *fmt* or *folD* mutants are deficient with respect to rates of translation initiation and growth. Interestingly, in *Salmonella enterica* serovar Typhimurium, gene

amplification compensates for these deficiencies. In several independent actinonin-resistant mutants, growth rates increase due to amplification of a chromosomal region encompassing *metZW*, identical tandemly repeated genes encoding tRNAi (Nilsson *et al*., 2006). Although the unformylated tRNAi is not the preferred substrate for initiation, increased intracellular concentrations of this RNA enhance the likelihood that it will abnormally bind the ribosomal P site to initiate translation. Eight mutants were characterized in which the fitness of an *fmt* or *folD* mutation was enhanced by 5 to 40 copies of a *metZW* amplicon, which varied in size from 1.9 to 94 kb. This amplification increases cellular levels of Met-tRNAi (Nilsson *et al*., 2006).

Despite amplification, these actinonin-resistant mutants were less fit than the wild-type strain in competition experiments conducted in mice (Nilsson *et al*., 2006). Typically, antibiotic resistance decreases bacterial fitness in the absence of selection. A drug-resistant phenotype frequently correlates with reduced growth rates, as discussed in an interesting review and references therein (Andersson, 2006). These observations might bode well medically for preventing the formation of persistent and dangerous antibiotic-resistant bacterial populations. However, as illustrated for resistance to actinonin, the risk for developing drug-resistant populations also depends on the frequency and effects of compensatory mutations that arise. Preference should be given to drugs where the fitness costs are high for the types of resistance that develop and where there is low probability for the acquisition of compensatory mutations (Andersson, 2006). Thus, it is important to consider fitness costs in the rational development of antibiotic treatments.

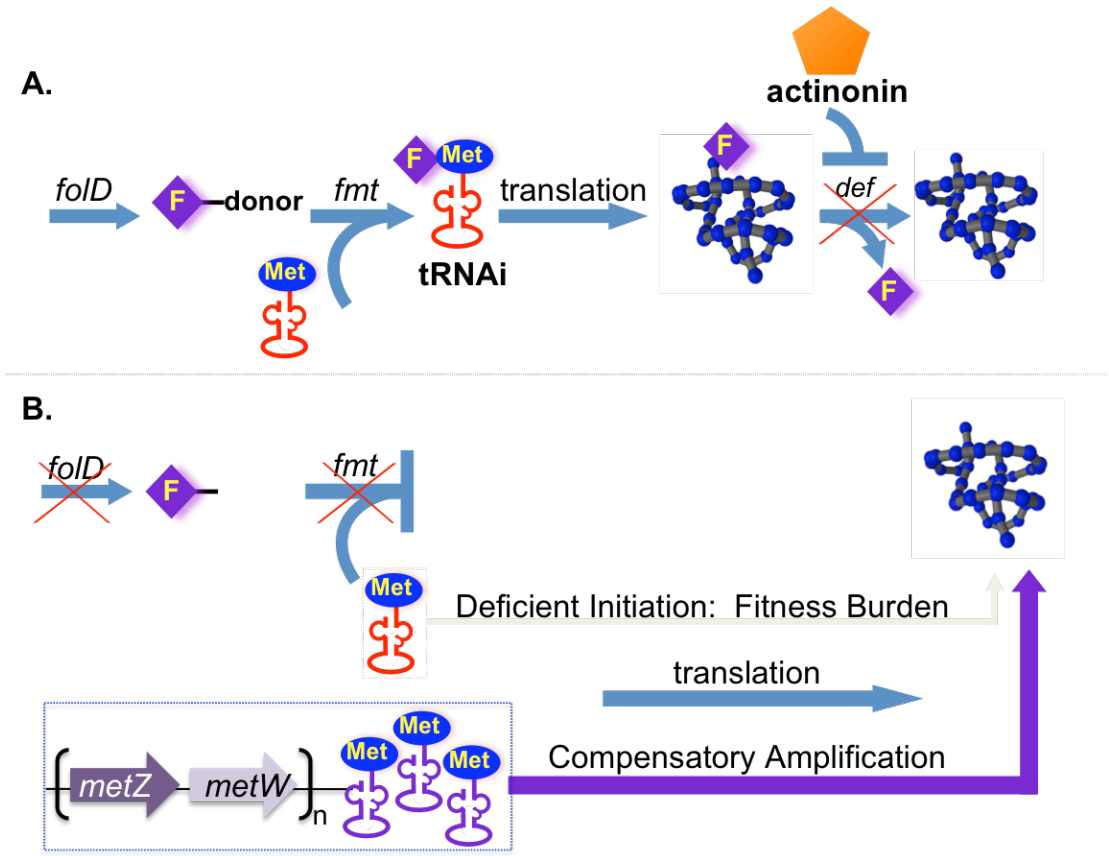


Figure 4.2: Gene amplification affects resistance to actinonin, an antibiotic that inhibits peptide deformylase, encoded by the *def* gene. (A) The toxic accumulation of formylated proteins has a bacteriostatic effect when actinonin prevents removal of the F group from N-terminal methionine residues. This group is incorporated into nascent bacterial polypeptides during translation initiation with formylated methionyl-tRNA (tRNAi). **(B)** Actinonin-resistant mutants survive by blocking the formylation of proteins via mutations in *folD*, encoding methylene-tetrahydrafolate dehydrogenase, or *fmt*, encoding formyltransferase. However, without formylated methionyl-tRNA, translation initiation and growth is severely impaired. Multiple copies of *metZW*, encoding tRNAi, compensate for the lack of formylation by increasing cellular concentrations of tRNAi. Improved translation initiation in amplification mutants correlates with increased fitness (Nilsson *et al*., 2006). F: Formyl; Met: Methionine.

In contrast to the example described above, reduced susceptibility to another peptide deformylase inhibitor (LBM415) results directly from increased amounts of the drug-target protein in *H. influenzae* (Dean *et al.*, 2007). LBM415 blocks the *def*-encoded enzyme, as does actinonin (Figure 4.2). Tandem amplification of the chromosomal *def* gene, with approximately 30-50 copies of the gene, increases the MIC of the drug 16 fold relative to the parent strain. Interestingly, *H. influenzae* sensitivity to LBM415 is also affected by the AcrAB-TolC efflux pump (Dean *et al.*, 2005). As noted earlier, the *acrAB* genes can increase drug resistance in *E. coli* by undergoing gene amplification (Nicoloff *et al*., 2007, Nicoloff *et al*., 2006).

Adaptation by amplification

Amplification regulates DNA and protein dosage without requiring permanent change. In this way, cells adapt to changing environmental conditions. Furthermore, when different genes are needed for a trait, maintaining these genes in close proximity allows them to be co-amplified. This arrangement can be selectively advantageous, and amplification serves as a driving force in the evolution of functionally-related gene clusters, such as pathogenicity islands (Reams & Neidle, 2004b). Nevertheless, amplified DNA is unstable and prone to deletion by homologous recombination. Amplified DNA increases the opportunities for mutations to accrue, and unnecessary or deleterious genes can be co-amplified with a beneficial segment. When selective pressures are removed or relaxed, amplified DNA tends not to be retained.

When selective pressure persists, permanent genetic changes may occur in amplified DNA that alleviate the need for amplification (Kugelberg *et al.*, 2006, Roth *et*

al., 2006). Mutation within a tandem array can generate a novel, beneficial allele that is selected and preserved within the population. The importance of duplication in generating paralogs has long been known. It is now becoming appreciated that the positive selection for multi-copy DNA is an important feature in the evolutionary process (Francino, 2005, Roth *et al*., 2006, Nilsson *et al*., 2006). Many copies of genes that encode proteins with weak or partial function contribute to the selection of variants with new or improved functions (Francino, 2005).

Mutations that arise under selection have led to controversial arguments concerning stress-induced mutability, alterations in mutation rates, and the ability of mutations to occur in non-growing cells (Roth *et al*., 2006, Slack *et al.*, 2006, Hastings *et al.*, 2004, Ponder *et al.*, 2005). Surprisingly, gene amplification lies at the heart of this controversy. More than twenty years ago, the *lac* genes of *E. coli* were developed as a model system for studying gene amplification (Tlsty *et al.*, 1984). Ongoing studies of *lac* mutants suggest that in populations of growth-limited cells, reversible increases and decreases in gene copy number may occur stochastically and affect the growth rate of subpopulations of cells with different levels of gene amplification. The calculation of mutation rates is complicated by the need to account for the number of genes, the number of cells, and the variable growth rates of individual cells and subpopulations during selection. Gene amplification has been proposed to account for mutation frequencies, which appear to increase during starvation, without invoking stress-induced alterations in mutation rates (Kugelberg *et al*., 2006, Roth *et al*., 2006). However, this explanation continues to be challenged (Stumpf *et al.*, 2007).

Tandem nucleotide repeats contribute to antigenic variation and molecular epidemiology

While the previous examples illustrate the effects of duplicating entire genes, short nucleotide repeats also contribute to important duplication events. Short tandem repeats, which affect gene regulation and protein production, can result from misalignment during DNA replication (Lovett, 2004). At some loci, allelic variation via the gain or loss of a discrete number of tandem-repeat units occurs as an important means of generating diversity.

Examples of tandem repeats involved in antigenic variation, drug resistance, and strain typing

Spontaneous sequence duplication controls capsular phase variation during pneumococcal infections. In *S. pneumoniae*, the polysaccharide capsule confers resistance to complement-mediated opsonization and phagocytosis. However, this capsule also inhibits adherence to and invasion of some eukaryotic cells. The ability to switch from a capsular to an acapsular form may facilitate the bacterial transition from carriage to invasive disease (Waite *et al.*, 2003). This phenotypic switch is triggered by spontaneous tandem duplication, which turns capsule production off. Precise excision of the duplication restores capsule production. Duplication-dependent variation has been observed in several different *S. pneumoniae* capsular serotypes, despite differences in the complexity and organization of the capsule loci (Waite *et al*., 2003, Waite *et al.*, 2001). In serotype 3, seven different perfect tandem duplications within the *cap3A* coding sequence can disrupt capsule production (Waite *et al*., 2001). Similar effects occur for a

perfect 22-bp tandem duplication, and its precise deletion, within the *tts* gene of serotype 3 (Waite *et al*., 2003). In serotype 8, disruption and restoration of the *cap8E* gene is mediated by a perfect 223 bp tandem duplication (Waite *et al*., 2003).

In a serotype 3 clinical isolate of *S. pneumoniae,* duplication causes macrolide resistance (Musher *et al.*, 2002). Such resistance was detected in a mutant that arose during intravenous drug therapy and resulted in patient death. Macrolides inhibit bacterial translation by binding to the 23S ribosomal subunit. In this example, the resistant mutant contains an 18-bp tandem repeat within the gene encoding ribosomal protein L22. The duplication inserts six amino acids adjacent to the proposed macrolide-binding site on the 23S rRNA and confers resistance to a variety of different macrolide antibiotics (Musher *et al*., 2002). As illustrated by these examples, tandem duplication can impact the progression and treatment of pneumococcal infections.

In other examples, repetitive DNA within a gene affects the immune-based clearance of pathogens. One gene with an extraordinary number of DNA repeats is *cagY* from *Helicobacter pylori*. This approximately 5.5 kb gene, which encodes a surface-exposed region of the type IV secretion system pilus, was examined in three different *H. pylori* strains. Direct repeats comprised of 16 or more nucleotides occur within this gene an average of 264 times (Aras *et al.*, 2003a). Assuming that recombination occurs randomly between each set of repeats, in-frame protein products would be expected in only 33% of the duplications. Amazingly, all 1215 potential recombination events in the three strains produce an in-frame CagY. Strains of *H. pylori* from infected patients as well as those that are passaged in the laboratory contain a heterogeneous population of expressed CagY proteins. This constantly diversifying pool of surface-exposed CagY

appears to evade detection by the humoral immune system. Despite cases involving years of *H. pylori* colonization, CagY elicits a minimal IgG response (Aras *et al*., 2003a).

In surface-exposed *Mycoplasma* proteins, DNA repeats are similarly employed for antigenic variation. One example involves the variable surface antigens (Vsa) of *M. pulmonis* (Simmons *et al.*, 2004, Simmons & Dybvig, 2003). A site-specific DNA inversion mechanism combines the *vsa* expression site with one of seven different *vsa* genes. Additional Vsa size alteration stems from an extensive tandem repeat region at the 3' end of the *vsa* genes. The number of tandem repeats changes when slipped-strand mispairing occurs during replication, thereby varying the size of these surface lipoproteins. Regardless of the type of the Vsa protein (i.e. which gene is expressed), cells encoding long proteins with many tandem repeats tend to be deficient in adherence and to resist lysis by complement (Simmons *et al*., 2004, Simmons & Dybvig, 2003). Furthermore, the number of tandem repeats in Vsa proteins correlates with other properties that affect pathogenicity. When Vsa proteins have few repeats, biofilm formation is favored. In contrast, microcolonies tend to form when Vsa proteins have many repeats (Simmons *et al.*, 2006). Stochastic changes in the size of Vsa proteins, resulting from variable number tandem repeats (VNTR), affect important mycoplasma virulence determinants.

__M__ultiple-__L__ocus __V__NTR __A__nalysis (MLVA) as an epidemiological tool for molecular differentiation of bacterial strains

There are many VNTR loci in mycoplasma species. For example, *Mycoplasma hyopneumoniae* has 22 such loci that affect the size of 12 proteins, including known

adhesins (de Castro *et al.*, 2006). Variation in the number of tandem repeats is relatively stable, at least in laboratory culture, and the size of a DNA fragment encompassing all repeats for a locus tends to be strain specific. Variations in the size of comparable loci in different strains will occur as discrete units corresponding to the number of repeats. Therefore, a PCR-based analysis of VNTRs in *M. hyopneumoniae* can help in strain typing and diagnosis (de Castro *et al*., 2006). A method for assessing multiple VNTR loci, termed MLVA, detects genetic polymorphisms within bacterial pathogens, as recently reviewed (Lindstedt, 2005, van Belkum, 2007). The applications for MLVA in molecular epidemiology are increasing, and a new database describing VNTR loci is available (Chang *et al.*, 2007). Systematic studies of the mutation rates of 28 individual VNTR loci in *Escherichia coli* O157:H7 and 43 loci in *Yersinia pestis* contribute to a better understanding of these hyper-variable regions (Vogler *et al.*, 2006, Vogler *et al.*, 2007).

Several clinical tests are examining the limits of the method. For example, it is important to understand how tandem repeats vary *in vivo*. Stability of the *spaA* gene of *Staphylococcus aureus*, which contains a highly variable repeat region, was tested during a 75-month longitudinal study (Kahl *et al.*, 2005). Samples were cultured from ten cystic fibrosis patients with persistent *S. aureus* infection, and 16 different *spa* types were identified in 142 isolates. During the course of the study, 8 independent mutational events were detected including deletions, duplications, and point mutations (Kahl *et al*., 2005). In other investigations, MLVA is being compared to different typing methods. For example, one study of hospital-acquired vancomycin-resistant eteroccocci suggested that

a combination of MLVA and multi-locus sequence typing might alleviate some problems with alternative methods of hospital epidemiology (Abele-Horn *et al.*, 2006).

Evolution of pathogens: a long history of gene duplication and amplification

VNTRs and other repetitive sequences contribute to dynamic rearrangements with adaptive and evolutionary significance. One central tenet of evolutionary models is that innovative functions develop from gene duplication and divergence, as reviewed and updated by Zhang (Zhang, 2003). Initially, it was postulated that the early stage after gene duplication was one of neutral selection. However, according to the adaptive radiation model, a period of positive selection for gene amplification precedes the selection of a variant allele (Francino, 2005). Consistent with this model, gene amplification occurs frequently and facilitates the adaptation to new environments (Reams & Neidle, 2004b). Recent studies address this process by investigating the emergence of new alleles from populations of amplified DNA segments (Kugelberg *et al*., 2006).

One study of 106 bacterial genomes, assesses the distribution of paralogs (i.e. homologous genes created by duplication after speciation) (Gevers *et al.*, 2004). In this analysis, an average of 24% of the protein-coding genes in each genome appear to have at least one paralog in the genome. Furthermore, the number of retained duplicates correlates with function. Preserved paralogs tend to be involved in adaptive processes, such as transcriptional regulation, transport, and organism-specific metabolism (Gevers *et al*., 2004). Another study of microbial genomes similarly concludes that paralog retention correlates with weak or ancillary gene function (Hooper & Berg, 2003). Interestingly,

clustered genes are associated with ancillary functions and frequently confer selectable benefit via gene amplification (Reams & Neidle, 2004b). In the bacterial genomes studied, most paralogs form a single unit rather than a block with neighboring genes that were co-duplicated (Gevers *et al*., 2004). However, the early duplications could involve larger segments that subsequently undergo substantial rearrangement. A recent study suggests that large amplicons readily acquire deletions (Kugelberg *et al*., 2006).

Examples of genome plasticity in bacterial pathogens: Mycobacterium *and* Francisella

In *Mycobacterium tuberculosis*, the causative agent of tuberculosis, there are two large multigene families that account for approximately 10% of the predicted coding sequences. While the functions of these families are unclear, they appear to affect pathogenesis. These PE and PPE gene families, named for characteristic patterns of proline (P) and glutamate (E) residues, are only observed in the genus *Mycobacterium.* A study of the expansion of these gene families in *M. tuberculosis* reveals a remarkable evolutionary history (Gey van Pittius *et al.*, 2006). A PE/PPE gene pair appears initially to have integrated into an *esx* gene cluster, a region of immunopathogenic importance that includes numerous genes (Gey van Pittius *et al*., 2006). Expansion of the PE/PPE families correlates with duplications of the entire clustered *esx* region. Subsequently, secondary subduplications of the PE and PPE genes occurred. One particular *esx* cluster was especially prone to generating PE and PPE expansions. From these duplications, two subfamilies emerged, PE_PGRS and PPE_MPTR (Gey van Pittius *et al*., 2006). These subfamilies have variable copies of repeated oligonucleotides (VNTR) that most likely

contribute to antigenic variation. Thus, several types of genetic duplication affect the evolution and function of these paralogs.

The dynamic nature of gene amplification makes it difficult to document transient genetic change. However, genome sequences occasionally capture interesting duplications. For example, a 34-kb region is duplicated in the 1.9 Mb chromosome of *Francisella tularensis*, an intracellular bacterium that causes tularemia (Larsson *et al.*, 2005). Interest in this organism is increasing because of its potential use as a bio-weapon. The duplicated region, considered to be a pathogenicity island, carries 25 genes, including those involved in virulence (de Bruin *et al.*, 2007, Lenco *et al.*, 2007). While the effect of duplication on virulence remains unclear, both copies of the duplicated island are flanked on one side by rRNA genes and on the other by an IS element. Thus, repetitive DNA may be involved in the genetic rearrangements (Larsson *et al*., 2005).

Duplicated chromosomal regions are not limited to pathogens. For example, *Dehalococcoides ethogenes*, a bacterium that dechlorinates groundwater pollutants, carries a 31-kb duplicated region on its 1.5 Mb chromosome (Seshadri *et al.*, 2005). This duplication appears to have been generated by non-homologous illegitimate recombination. Clearly, diverse mechanisms promote rearrangements and genetic flexibility.

Patterns of repetitive DNA: Mycoplasma *spp.,* H. pylori*,* Ehrlichia ruminantium*, and* Phytoplasma *strains*

Various types of DNA repeats contribute to rearrangements in bacterial genomes. Several studies of repetitive DNA focus on *Mycoplasma*, bacteria with small genomes

(0.58 to 1.35 Mb) and a striking proportion of repeated DNA (Mrazek, 2006, Papazisi *et al.*, 2003, Rocha & Blanchard, 2002). These obligate parasites, which cause disease in humans and other vertebrates, are interesting subjects of genomic analysis, as described in recent papers and references therein (Mrazek, 2006, Papazisi *et al*., 2003, Rocha & Blanchard, 2002). Some repeats are involved in high-frequency antigenic variation. The importance of this process is exemplified by *M. gallisepticum*, for which an estimated 16% of the genome serves as a reservoir for the variation of a single antigen (Papazisi *et al*., 2003). In some cases, protein variation depends on simple sequence repeats, tandem iterations of a single nucleotide or short oligonucleotide. Such repeats, which are typically underrepresented in prokaryotic genomes, are common in *Mycoplasma*. Protein variation may result from these repeats within coding sequences that cause integral changes in the number of codons or from such repeats in regulatory regions that control gene expression. However, a systematic analysis of simple sequence repeats reveals that their distribution and composition are characteristic of diverse functions that extend beyond their roles in antigenic variation (Mrazek, 2006).

Another bacterium with extensive non-randomly distributed chromosomal repeats is *H. pylori*. As the only organism that colonizes the highly acidic human stomach, this pathogen must diversify its genetic material in the absence of cohabitating species (Aras *et al.*, 2003b). The observed diversity in *H. pylori* appears to stem from extensive direct repeats that generate dynamic genomic rearrangements within individual hosts (Aras *et al*., 2003b). Evidence is presented that programmed recombination hotspots mediate hyper-variation at specific loci (Aras *et al*., 2003a, Aras *et al*., 2003b). Dynamic rearrangements are also observed in the genome of *Ehrlichia ruminantium*, an obligate

intracellular bacterium that causes heartwater, a disease of ruminants (Collins *et al.*, 2005). In this example, the rearrangements involve a large number of duplicated and tandemly repeated sequences, some of which exhibit variable copy number (Collins *et al*., 2005). This pattern leads to an atypically small amount of DNA being used for protein-coding sequence, 62% of the genome. This percentage is remarkable considering the minimal genome size, 1.5 Mb. Another atypical pattern, in a phytoplasma genome, involves multiple clusters of diverse, functionally unrelated genes. This type of mosaic structure may reflect frequent duplication, deletion and rearrangements mediated by repetitive extragenic sequence elements (Jomantiene & Davis, 2006).

Conclusions and Future Perspective

Gene duplication and amplification can have significant medical consequences. However, its dynamic nature makes amplification difficult to identify, and this type of genetic variation often goes undetected. Unstable phenotypes are commonly attributed to the loss of plasmids or mobile elements. However, instability may instead result from the loss of amplification after selective conditions are removed or relaxed. As advancements in molecular methods improve genetic characterization and DNA sequencing, novel examples of genetic duplication and amplification continue to emerge. Furthermore, computational approaches to genome analyses make it clear that significant patterns of repetitive DNA contribute to frequent chromosomal rearrangements.

Gene amplification is only one of many factors affecting increased antibiotic resistance in bacteria. Nevertheless, the profound clinical importance of such resistance is a reminder of the ongoing and exigent challenges in treating life-threatening illnesses

with antibiotics. Recent studies highlight the roles that amplification can play in augmenting resistance, facilitating the adaptation of bacteria to survive new drug regimens, and ameliorating the fitness costs of other mutations (Francia & Clewell, 2002, Nicoloff *et al*., 2007, Nilsson *et al*., 2006, Seoane *et al*., 2003, Yanai *et al*., 2006). It is important to understand not only the detailed molecular basis of drug resistance but also the fitness costs associated with genetic change (Andersson, 2006). These factors help determine how long drug-resistant bacterial populations will persist after the use of an antibiotic has been reduced or eliminated.

While gene amplification is typically a reversible process, increased gene dosage enhances the opportunity for mutations to accrue. If a novel allele on an amplified fragment provides a selectable benefit, it may be retained and enable the copy number of the fragment to decrease (Kugelberg *et al*., 2006, Roth *et al*., 2006). This process has important ramifications for the evolution of new gene functions and for the retention of paralogs during evolution (Francino, 2005, Gevers *et al*., 2004, Reams & Neidle, 2004b). In some cases, the selectable advantage of increased gene dosage is stabilized by mutations that force the duplicated region to be retained in more than one copy (Cerquetti *et al*., 2005, Cerquetti *et al*., 2006b, Kroll *et al*., 1991). For example, such stabilization is mediated by a deletion in a portion of the duplicated *cap* locus in *H. influenzae* (Kroll *et al*., 1991).

The amplification of the *H. influenzae cap* locus, which can cause severe disease and contribute to vaccine failure, highlights the importance of monitoring the copy number of this locus in circulating strains. In general, renewed awareness and attention to the copy number of duplicated segments in bacterial strains is warranted. In some cases,

molecular epidemiology is even facilitated by variation in the number of tandem repeats (Lindstedt, 2005, van Belkum, 2007). It is becoming increasingly apparent that bacteria use repetitive DNA sequences such as variable number tandem repeats as a way of diversifying their surface antigens, e.g. (Mrazek, 2006, Papazisi *et al*., 2003, Rocha & Blanchard, 2002). Antigenic variation often occurs at rates faster than the immune system can respond. Therefore, genetic variability offers serious challenges to immunological clearance and vaccine-dependent prevention of bacterial infections. Bacterial-host interactions often involve adaptive responses between organisms, and genetic duplication and amplification can facilitate such adaptation.

Many aspects of the diverse mechanisms affecting gene amplification remain poorly understood, and few model systems have been developed to investigate gene duplication in a systematic fashion. An exciting area of future research focuses on the mechanisms of gene amplification and factors that affect its occurrence. The widespread availability of genomic DNA sequence is associated with increased interest in assessing the transcripts and proteins of individual bacteria and communities of organisms. Few of these transcriptome and proteome analyses address changes that occur in gene dosage. Although the sensitivity of many methods makes it difficult to study infrequent and transient changes in DNA copy number, the judicious design and modification of experimental protocols should allow some amplification events to be studied in detail. Techniques typically used for transcriptional analysis can be applied to studies of copy-number effects at the level of DNA. For example, DNA microarray methods and real-time quantitative PCR can be used to assess the prevalence and location of some genomic duplications. Ongoing research to characterize genetic duplication and amplification has

the potential to improve the rational use and development of effective antibiotics, vaccines, and chemotherapeutic treatments.

CHAPTER 5

A SURVEY OF GENETIC DUPLICATION AND AMPLIFICATION IN THE *ACINETOBACTER BAYLYI* ADP1 GENOME[4]

[4] Craven, S.H., K.T Elliott and E.L. Neidle. To be submitted to *Journal of Bacteriology*.

Introduction

The widespread importance of prokaryotic gene amplification is underscored by diverse manifestations including antibiotic resistance (Nicoloff *et al.*, 2007, Nicoloff *et al.*, 2006), enhanced virulence, and immune evasion by pathogenic microorganisms (Simmons *et al.*, 2004, Waite *et al.*, 2003). Amplification also enables organisms to adapt to highly variable, limiting or even extreme environments (Reams & Neidle, 2003). In higher organisms, genetic duplication and amplification are linked to the resistance of *Plasmodium* to antimalarial chemotherapeutics (Valderramos & Fidock, 2006), and to acquired resistance to methotrexate and other therapeutics in the treatment of human cancer (Pepper *et al.* 2009). Amplification also plays a key role in oncogene activation during the development of eukaryotic cell cancers (reviewed in Albertson, 2006). Despite the pervasiveness of genetic amplification, the scope of this type of genome plasticity is not well understood. Here, we present the soil bacterium *Acinetobacter baylyi* ADP1 as a unique and tractable model system for genome-wide studies of duplication and amplification. Inherent advantages of the ADP1 system allow a broad survey of naturally occurring duplication sites.

Strain ADP1 is uniquely suited for studies of genetic duplication and amplification

During studies of aromatic compound catabolism in ADP1, gene amplification was identified as a mechanism for increasing levels of benzoate catabolism in strains lacking wild-type transcriptional activation (Reams and Neidle, 2003). A two-step model was proposed to describe amplification in these strains (Figure 5.1), in which the first creates a chromosomal duplication and the second generates additional copies of the

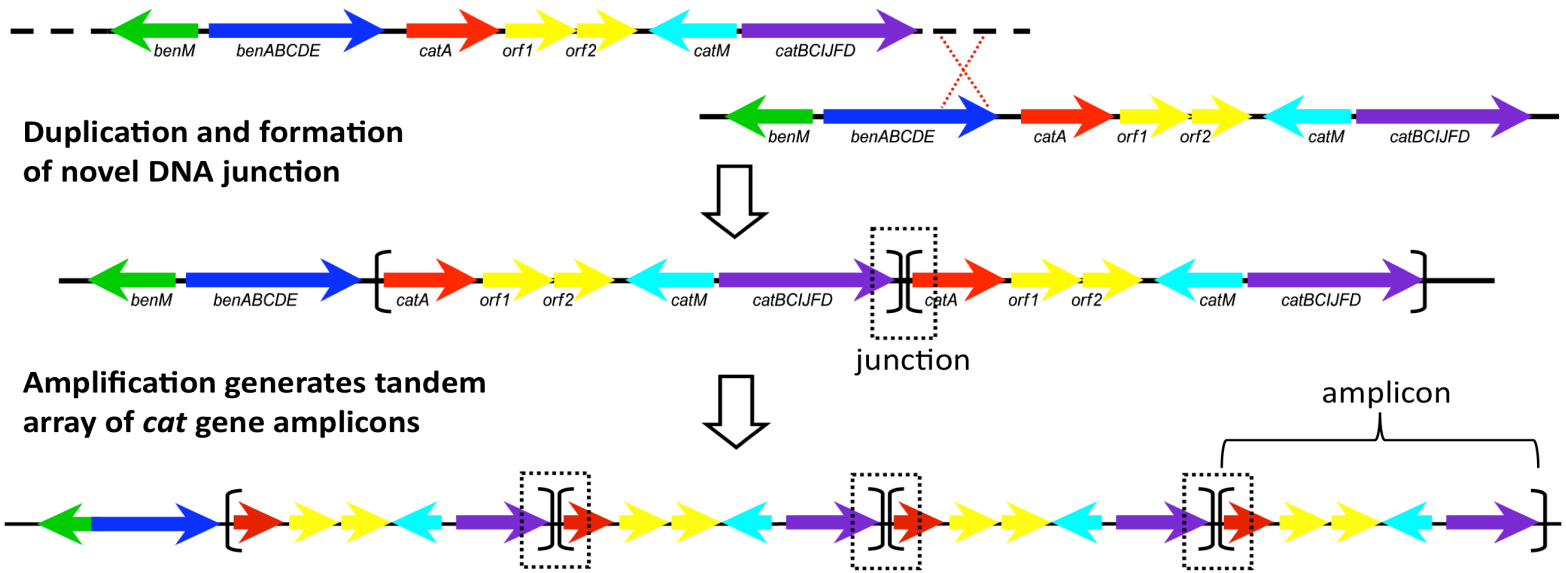


Figure 5.1: Model for genetic amplification of *cat* genes in *Acinetobacter baylyi* ADP1. Amplification occurs by a two-step process. In the first step, a recombination event fuses DNA normally downstream of *catD* with DNA normally upstream of *catA*. Duplication generates a characteristic novel DNA junction (black box). Following formation of the duplication, further expansion generates a tandem head-to-tail array of the entire duplicated region, termed an amplicon.

duplicated region, or amplicon (Reams and Neidle, 2004a). Based on these results, an experimental system was developed to characterize the DNA sequences that undergo recombination to generate tandem chromosomal duplications and allow further amplification (Reams and Neidle, 2004a). The ADP1 system offers non-trivial advantages to overcome problems that commonly thwart studies of gene amplification in other organisms. Specifically, 1) the selection method yields exclusively mutants with genetic amplifications and does not select for simple point mutations. 2) The genomic rearrangements are large (>13 kbp), allowing direct visualization of chromosomal changes using pulsed-field gel electrophoresis (PFGE). 3) The gene amplification is retained during growth with benzoate as the sole source of carbon. This retention is important, as the reversibility of transient gene amplification is typically a challenge to laboratory study. Finally, and most importantly for ongoing studies, 4) a simple transformation assay allows the precise DNA sequence at the endpoints of the duplicated region to be determined by exploiting the unusually high efficiency of natural transformation and recombination in *A. baylyi* (Reams and Neidle, 2004a; Young *et al*., 2005).

To what extent does sequence identity contribute to recombination between distant DNA fragments in the spontaneous generation of chromosomal duplications?

It is well established that DNA sequence identity between different genetic regions promotes recombination events that create chromosomal duplications (Romero and Palacios, 1997). This identity may be extensive (>1kb) as is the case for duplication between insertion sequence elements (Sun *et al*., 2009, Haack & Roth, 1995, Kroll *et al.*,

1991) and ribosomal RNA operons (Anderson and Roth, 1981). When long stretches of DNA are involved, the bacterial RecA protein contributes to this process, known as homologous recombination. However, extensive DNA sequence identity is not required, and several RecA-independent processes, known as illegitimate recombination, also participate in duplication formation (Ikeda *et al*., 2004, Pitcher *et al.*, 2007).

Tandem duplications in bacteria can affect almost any chromosomal locus with frequencies that vary from 10^{-2} to 10^{-5} (Sun *et al*., 2009; Anderson and Roth 1981). The best-characterized methods of duplication formation include homologous recombination, site-specific illegitimate recombination, DNA slippage, and short-homology dependent illegitimate recombination (Ikeda *et al*., 2004, Romero and Palacios, 1997). The mechanisms associated with these processes all rely on some degree of sequence identity between the different DNA regions involved in duplication formation. Distant regions of genomic DNA can also be joined together in novel combinations by non-homologous end joining (NHEJ) (Ikeda *et al*., 2004). While it was once thought that NHEJ occurs only in eukaryotes, recent studies demonstrate its occurrence in some bacteria (Bowater & Doherty, 2006). However, homologs of known proteins mediating NHEJ appear absent in *A. baylyi* ADP1.

In previous studies, sequences of precise duplications were characterized in approximately 100 independent *A. baylyi* amplification mutants. Surprisingly few of the rearrangements involved long or short regions of repetitive DNA sequences (Reams and Neidle, 2004a). Interestingly, some of the illegitimate recombination events display atypical position specificity. Six different duplication sites were each identical in two to six independently isolated mutants (Reams and Neidle, 2004a). These duplications occur

spontaneously in the absence of RecA and appear to form by a novel process, now termed Position-Specific Illegitimate Recombination (PSIR).

As described in this chapter, the ADP1 model system was expanded to allow a comprehensive survey of duplication sites that occur throughout the genome. One long-term goal is to test whether repetitive DNA sequences play a more significant role in duplication formation in some regions of the chromosome than in others. This approach will allow us to determine whether the spectrum and frequency of different types of rearrangements, including the unusual PSIR events, vary according to chromosomal position. No such systematic study has yet been conducted in any organism.

Results

Generation of an A. baylyi *strain (ACN854) to isolate new amplification mutants*

A parent strain, designated ACN854, was engineered to allow the selection and characterization of novel genome rearrangements. While this new strain differed in important ways from the strain used in previous amplification studies (ACN293, Figure 5.2A), both strains similarly lacked functional copies of *catM* and *benM.* Inactivation of these genes, which encode transcriptional activators typically needed for benzoate consumption, prevents the parent strain from growing on benzoate (Ben^-). However, mutants that acquire the ability to consume benzoate (Ben^+) are readily selected on solid medium (Reams and Neidle, 2003). The ACN293-derived Ben^+ mutants increase the expression of requisite genes via a tandem head-to-tail array of multiple copies of a

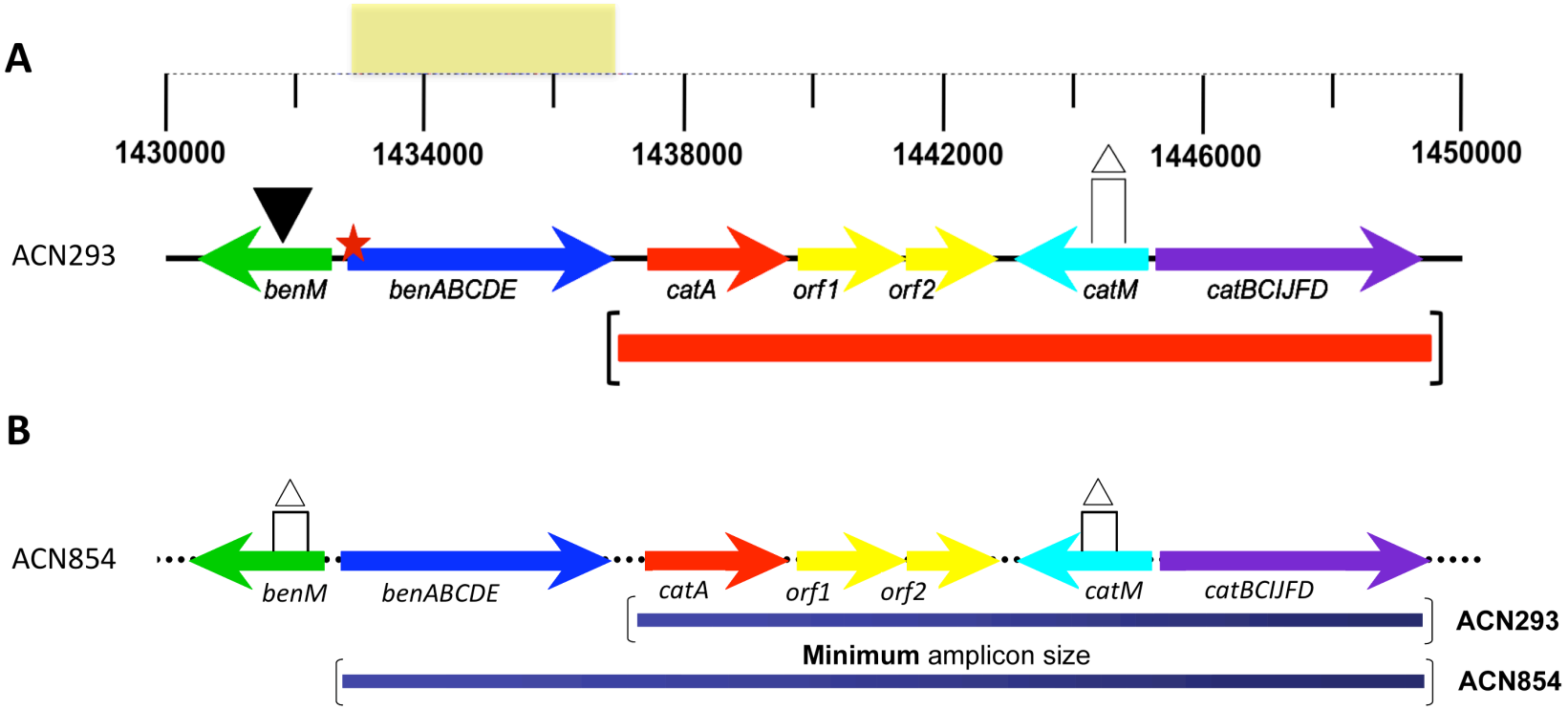


Figure 5.2: Schematic showing the genotype and relevant characteristics of strains ACN293 and ACN854. In panel A, the *ben-cat* region of ACN293 is shown. *benM* is disrupted with an antibiotic cassette (black triangle), *catM* contains a partial deletion (Δ), and a the *benA* promoter mutation is indicated (red star). The top line indicates approximate location within the ADP1 genome (GenBank entry CR543861). The yellow box delineates the ~5-kbp region in which all previously characterized upstream amplicon endpoints were found. Panel B highlights the genotype of newly engineered strain ACN854, with deletions in both *benM* and *catM* (Δ). Note that for growth on benzoate, the minimum amplicon size must span the entire *ben-cat* region in ACN854-derived strains.

chromosomal region, termed the amplicon. While the size of the amplicon can vary in different Ben$^+$ mutants, it always encompasses two distinct but neighboring transcriptional units, *catA* and the *catBCIJFD* operon. Basal expression from multiple copies of these seven *cat* genes increases the production of enzymes that convert catechol to tricarboxylic acid cycle intermediates.

In ACN293 and its derivatives, a point mutation in the *benA* promoter causes high-level *ben*-gene expression without the normally required BenM transcriptional activator. Under the original selection devised by Reams and Neidle, the *benA* promoter mutation is not tolerated in multiple copy and thereby restricts the upstream amplicon endpoint to an approximately 5 kbp region between the *benA* and *catA* promoters (gold box, Figure 5.2A) (Reams and Neidle, 2004a). Over-expression of the *ben* genes relative to the *cat* genes could create a metabolic imbalance and accumulation of the toxic intermediate catechol (See Figure 3.1 in Chapter 3). Thus, the use of ACN293 to isolate Ben$^+$ mutants may limit the types of genetic rearrangements that can be characterized.

Strain ACN854 was engineered to carry the wild-type *benA* promoter without the point mutation that confers high-level *ben*-gene expression. As expected, this strain lacking CatM and BenM failed to form colonies or grow in liquid culture with benzoate as the sole carbon source (data not shown). To enable growth on benzoate in this strain, an amplicon should need to span more than 16 kbp to encompass the *benABCDE* operon as well as *catA* and the *catBCIJFD* operon (Figure 5.2B). The importance of the ACN854 parent compared to the previously used ACN293 is two-fold: 1) growth on benzoate should require amplification of a larger genomic region (>16 kbp versus 10 kbp) and 2)

to generate the desired Ben$^+$ phenotype, a recombination event should occur upstream of *benA*, in a chromosomal region not characterized in previous studies.

Selection and genomic characterization of ACN854-derived Ben+ mutants

To select potential amplification mutants, single ACN854 colonies were grown to high cell density and plated (~10^9 cells) on solid benzoate medium. Selective plates were incubated at 30°C for 7-14 days until colonies were visible. During this time, approximately 1 colony formed for every 10^8 cells plated. A single colony was selected from each plate and considered an independent isolate. In total, twenty independent colonies were streak purified to obtain individual mutants for further characterization. Strains described in this chapter are listed in Table 5.1. In each of these Ben$^+$ strains, PCR analysis confirmed the deletions within *benM* and *catM*, consistent with retention of the parental alleles.

To assess the chromosomal configuration of the Ben$^+$ mutants relative to the ACN854 parent, total genomic DNA was digested with a diagnostic restriction enzyme (NotI) and separated by pulsed-field gel electrophoresis (PFGE). Digestion of genomic DNA from the wild type and parent strain (ACN854) with NotI generates a representative pattern with 6 chromosomal fragments ranging in size from 1380 to 97 kbp (Gralton *et al.*, 1997). In mutants harboring amplification of the *ben-cat* region, gross chromosomal rearrangements are visible by PFGE (Reams and Neidle, 2003). All twenty of the ACN854-derived Ben$^+$ mutants analyzed here displayed chromosomal patterns that were consistent with the presence of gene amplification (data not shown).

Table 5.1: Strains and plasmids used in this study

Strains used in this study		
Strain	**Relevant characteristics**[a]	**Source**
ADP1	Wild-type (BD413)	(Juni, 1969)
ACN293	*benM*ΩS4036, *benA*5147, Δ*catM*5293; benzoate$^-$ parent from which mutants were selected	(Reams, 2004)
ACN389	Δ*benM*5389 in wild-type ADP1 background	(Bundy *et al.*, 2002)
ACN871	Ant$^+$ mutant derived from ACN293; harbors an ~270 kbp amplicon encompassing *cat* genes	This study
ACN872	Ant$^+$ mutant derived from ACN293; harbors an ~193 kbp amplicon encompassing *cat* genes	This study
ACN873	Ant$^+$ mutant derived from ACN293; harbors an ~18 kbp amplicon encompassing *cat* genes	This study
ACN874	Ant$^+$ mutant derived from ACN293; harbors an ~77 kbp amplicon encompassing *cat* genes	This study
ACN875	Ant$^+$ mutant derived from ACN293; harbors an ~102 kbp amplicon encompassing *cat* genes	This study
ACN877	Ant$^+$ mutant derived from ACN293; harbors an ~264 kbp amplicon encompassing *cat* genes	This study
ACN883	Ant$^+$ mutant derived from ACN293; harbors an ~273 kbp amplicon encompassing *cat* genes	This study
ACN854	Δ*benM*5389, Δ*catM*5293; benzoate- parent from which mutants were selected	This study
ACN868	Ben$^+$ mutant derived from ACN854; harbors an ~35 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN869	Ben$^+$ mutant derived from ACN854; harbors an ~35.5 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN896	Ben$^+$ mutant derived from ACN854; harbors an ~35 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN898	Ant$^+$ mutant derived from ACN854; harbors an ~27 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN899	Ben$^+$ mutant derived from ACN854; harbors an ~47 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN906	Ben$^+$ mutant derived from ACN854; harbors an ~35.5 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN950	Ben$^+$ mutant derived from ACN854; harbors an ~32 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN951	Ant$^+$ mutant derived from ACN854; harbors an ~34 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN956	Ben$^+$ mutant derived from ACN854; harbors an ~29 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN1007	Ben$^+$ mutant derived from ACN854; harbors an ~32 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN1118	Ben$^+$ mutant derived from ACN854; harbors an ~34 kbp amplicon encompassing the *ben* and *cat* genes	This study
ACN1123	Ben$^+$ mutant derived from ACN854; harbors an ~94 kbp amplicon	This study

	encompassing the *ben* and *cat* genes	
ACN1023	*benM*ΩS4036, *benA*5147, Δ*cat*5825 (*catA-catD*); complete *cat* region (1439430-1449722)[b] inserted into locus ACIAD3593	This study
ACN1024	*benM*ΩS4036, *benA*5147, Δ*cat*5825; complete *cat* region (1439430-1449722)[b] inserted into locus ACIAD3593 with *catM*ΩKm; Ben$^-$ parent from which mutants were selected	This study
ACN1026	Ben$^+$ mutant derived from ACN1024	This study
ACN1027	Ben$^+$ mutant derived from ACN1024	This study
ACN1028	Ben$^+$ mutant derived from ACN1024	This study
ACN1029	Ben$^+$ mutant derived from ACN1024	This study
ACN1032	Ben$^+$ mutant derived from ACN1024	This study
ACN1033	Ben$^+$ mutant derived from ACN1024	This study
ACN1036	*benM*ΩS4036, *benA*5147, Δ*cat*5825 (*catA-catD*); complete *cat* region (1439430-1449722)[b] inserted into *vanK* locus (ACIAD0982)	This study
ACN1039	*benM*ΩS4036, *benA*5147, Δ*cat*5825 (*catA-catD*); complete *cat* region (1439430-1449722)[b] inserted into *vanK* locus (ACIAD0982) with *catM*ΩKm; Ben$^-$ parent from which mutants were selected	This study
ACN1040	Ben$^+$ mutant derived from ACN1039	This study
ACN1041	Ben$^+$ mutant derived from ACN1039	This study
ACN1042	Ben$^+$ mutant derived from ACN1039	This study
ACN1043	Ben$^+$ mutant derived from ACN1039	This study
ACN1044	Ben$^+$ mutant derived from ACN1039	This study
Plasmids used in this study		
Plasmid	**Relevant characteristics**	**Source**
pZErO-2	Km^R; Zero background cloning vector	Invitrogen
pET-21b	Ap^R; cloning vector	Novagen
pCR2.1-TOPO	AprKmr; PCR cloning vector	Invitrogen
pBAC6	Ap^R, Km^R; ΩKm in *Hinc*II site of *catM*	(Ezezika *et al.* 2006)
pBAC841	Universal *cat* region shuttle plasmid; contains *cat* DNA (1439430-1449722)[b] flanked by unique restriction sites to facilitate cloning; See *Experimental Procedures* for details of plasmid engineering	This study
pBAC850	Shuttle plasmid designed to insert *catA-catD* into the ACIAD3593 chromosomal locus	This study
pBAC874	Shuttle plasmid designed to insert *catA-catD* into the ACIAD0982 chromosomal locus	This study
pBAC857	ACN868 junction plasmid: pZErO2 with *Eco*RI insert (1432131-1431778; 1466991-1464347)[b]	This study
pBAC858	ACN869 junction plasmid pZErO2 with *Eco*RI insert (1432131-1431838; 1467382-1464347)[b]	This study
pBAC859	ACN896 junction plasmid pZErO2 with *Eco*RI insert (1437229-1433352; 1468488-1464305)[b]	This study
pBAC860	ACN898 junction plasmid pZErO2 with *Hind*III insert (1437877-1434048; 1461133-1460612)[b]	This study
pBAC861	ACN899 junction plasmid	This study

	pZErO2 with *Hind*III insert (1430928-1428708; 1475753-1457359) [b]	
pBAC862	ACN950 junction plasmid pZErO2 with *Eco*RI insert (1464347-1465535; 1433380-1437229) [b]	This study
pBAC863	ACN951 junction plasmid pZErO2 with *Eco*RI insert (1450761-1455727; 1421717-1437229) [b]	This study
pBAC864	ACN956 junction plasmid pZErO2 with *Eco*RI insert (1437229-1432834; 1462267-1450763) [b]	This study
pBAC865	ACN906 junction plasmid pZErO2 with *Eco*RI insert (1432131-1431838; 1467382-1464347) [b]	This study
pBAC866	ACN1007 junction plasmid pZErO2 with *Eco*RI insert (1437229-1433380; 1465535-1464347) [b]	This study
pBAC867	ACN1118 junction plasmid pZErO2 with *Eco*RI insert (1450761-1455727; 1421717-1437229) [b]	This study
pBAC868	ACN1123 junction plasmid pZErO2 with *Eco*RI insert (1511262-1516254; 1421888-1432129) [b]	This study

[a]Km^R, kanamycin resistant; Ap^R, ampicillin resistant; ΩKm, omega cassette conferring Km^R; ΩS, omega cassette conferring resistance to streptomycin and spectinomycin.

[b]Position in the ADP1 genome sequence according to GenBank entry CR543861

Assessment of amplicon copy number using a quantitative PCR (qPCR) method

To confirm that chromosomal rearrangements visualized by PFGE correlated with the presence of multiple copies of the *ben-cat* region, quantitative PCR was used to estimate *catA* copy number. For copy number determination, mutant strains were first passaged on benzoate medium for approximately 100 generations to favor stable amplicon copy number within the cell population. A fluorescently labeled Taqman probe was designed to detect *catA*, and a second probe was designed to detect *antA*, a gene distantly located on the ADP1 chromosome. In populations of wild-type ADP1, *catA* and *antA* should each be present in single copy on the chromosome (*catA*:*antA* equal to 1). In amplification strains, the *catA* gene will be present in each copy of the amplicon and the distant *antA* will remain in single copy and serve as an endogenous control (*catA*:*antA* greater than 1). The *catA:antA* ratio, therefore, is inferred to correlate with the amplicon copy number in individual mutants. Using this technique, amplicon copy was determined in ten independent strains. Table 5.2 reveals variation in amplicon copy number.

Determination of DNA sequences at the precise position of the duplication.

Duplication formation is marked by a novel DNA rearrangement in which normally distant chromosomal regions are fused. This juxtaposition generates a novel DNA sequence referred to as a junction (JCN) at the endpoints of each amplicon (black box, Figure 5.1). Here, a transformation assay was used to determine the precise DNA sequence at duplication junctions in selected mutants (developed by Reams and Neidle,

Table 5.2: Amplicon size and copy number in ACN854-derived strains.

JCN Identifier[a]	Strain	Amplicon Size[b] (kbp)	Amplicon Copy Number[c]
JCN868	ACN868	35.2	10
JCN869	ACN869	35.5	8
	ACN906	35.5	10
JCN896	ACN896	35	10
JCN898	ACN898	26.9	33
JCN899	ACN899	47	7
JCN950	ACN950	32.1	7
	ACN1007	32.1	5
JCN951	ACN951	34	6
	ACN1118	34	not determined
JCN956	ACN956	29.4	4
JCN1123	ACN1123	94	not determined

[a]Junctions are identified by the strain number in which they were first described. Note that some junctions were identified in multiple strains.

[b]Amplicon size is determined using the DNA sequence of the junction and the known sequence of the ADP1 genome (GenBank entry: CR543861).

[c]Amplicon copy number determined in benzoate-grown populations using qPCR. Copy number was determined in triplicate samples and averages were rounded to the nearest whole number.

2004a and described in *Experimental Procedures*). Junction (JCN) sequences were determined in twelve ACN854-derived strains and are shown in Figure 5.3. These sequences allow us to infer the exact size of the amplicon, ranging from just under 27 kbp to more than 94 kbp, and in all cases agree with previous size estimates from PFGE analysis of these strains.

In addition to amplicon size, junction DNA offers information about the DNA aligned during recombination to generate the initial genetic duplication. Alignment of each junction with the wild-type ADP1 genome sequence reveals that sequence identity in the immediate vicinity of the recombination site is minimal. The number of identical nucleotides at the precise recombination site ranged from zero to four bases (bold and underlined in Figure 5.3) in characterized junctions. In all cases examined, sequence patterns were indicative of illegitimate rather than homologous recombination. Interestingly, despite the absence of notable homology, the identical junction sequences were isolated in independent mutants. JCN869, generating an amplicon 35.5 kbp in length, which contains only two identical nucleotides at the junction, was isolated in independent strains, ACN869 and ACN906. Two additional junctions were each isolated from two independent amplification mutants. These junctions, JCN950 and JCN951, contain zero and two base pairs of identity at the join point respectively. The independent recurrence of identical junctions resembles a previously identified type of illegitimate recombination in *A. baylyi* (Reams and Neidle, 2004a), termed PSIR.

JCN868
TTGCATTGTTTGGTATGATGTTT **GGTG** CCCTGATTTTTGGCACCATCGCC 1431750
TGCTGGCAAAGCCAAAATATGAA **GGTG** CCCTGATTTTTGGCACCATCGCC JCN
TGCTGGCAAAGCCAAAATATGAA **GGTG** GTCTGTTTGAGGGCGGTGCTGAT 1466960

JCN869 (2X)
GAACATTTGGGTGTGAGCCGTAAA **AA** AGTCATTGCGGTCTGTATTATTTT 1431809
AGATATTTCGTATCAGACAAAAGT **AA** AGTCATTGCGGTCTGTATTATTTT JCN
AGATATTTCGTATCAGACAAAAGT **AA** TGGATCTGAAAAAAGCGTCATTTG 1467352

JCN896
AGGCTTAGGCGAGCTTGGATAAAGC AAAATTTTTTCATCAATTAAATCAT 1433322
CAAAACCGACAAGACATGGACGCTT AAAATTTTTTCATCAATTAAATCAT JCN
CAAAACCGACAAGACATGGACGCTT GCATTTAAATGGCTATGGACAAAAG 1468459

JCN898
AGAAGGCAGGGGCTTGACCCATTA **A** ATGCTTTCTTCAATTTGGAAAATTG 1434019
AATACGATCACATCCCGCCTGAAT **A** ATGCTTTCTTCAATTTGGAAAATTG JCN
AATACGATCACATCCCGCCTGAAT **A** TTGCCCAGAGAAACAATATGCTGGC 1461104

JCN899
AACTGACTTTGAATCGGCTTATGA **CC** GCCTACACGATAAAAAACCTGATC 1428679
CAAACATTAGCGTCAAAAACTGCC **CC** GCCTACACGATAAAAAACCTGATC JCN
CAAACATTAGCGTCAAAAACTGCC **CC** TAAAGAGGTCGTGCAAAATAATTT 1475725

JCN950 (2X)
TTTTTTCATCAATTAAATCATTTA AGTGCACGCCTTTATCTTTCATTTGA 1433351
CATGTTGGTTGCGATCTTCCCAAT AGTGCACGCCTTTATCTTTCATTTGA JCN
CATGTTGGTTGCGATCTTCCCAAT TTTGACCAACACCACCTTGGGTTTAC 1465507

JCN951 (2X)
TAGTACACATTTGGAGTTTAAGCA **TC** TTGAATATGAAACTCTTTTTTTAG 1421688
CAGCCAGATCATGCATAAACCAGC **TC** TTGAATATGAAACTCTTTTTTTAG JCN
CAGCCAGATCATGCATAAACCAGC **TC** ATACCGCAGTCGTCAAAGTCAATC 1455697

JCN956
TAACTTTCTGGCACTGAGTATTCCT GCGGTTCTTGGTATTGTCACGGTAT 1432804
GAATGGTCTTATACGCAAGATTGGG GCGGTTCTTGGTATTGTCACGGTAT JCN
GAATGGTCTTATACGCAAGATTGGG TTGTACAAATGGTAAAATCTGGCTC 1462238

JCN1123
TCATGTATTAAAACTAATGGGGCA **T** CAGCGTAAAATTGCCTTGGAAGTGC 1421859
CTTAACATTAGATGTTTATCAAAT **T** CAGCGTAAAATTGCCTTGGAAGTGC JCN
CTTAACATTAGATGTTTATCAAAT **T** GATATTAAAAATAGAATTTTATTAT 1516225

Figure 5.3: DNA sequence involved in recombination to generate tandem duplication in mutants derived from ACN854. For each alignment, the middle line (black text) indicates the novel DNA junction (JCN) from an amplification strain. The top line (red text) represents the parental upstream sequence, with the first displayed nucleotide identified by its position in the ADP1 genome sequence (GenBank entry: CR543861). The bottom line is the parental downstream sequence, with position indicated. Identical nucleotides at the DNA junction are represented by bold and underlined text. Homology in the vicinity of the recombination junction is represented by underlined text. Three junctions, JCN869, JCN950, JCN951, were each identified in two independent amplification strains and are highlighted in bold text.

Position-specific DNA junctions can be detected in cell populations in the absence of selective pressure

Despite absence of sequence homology or apparent DNA sequence patterns, multiple recombination events show site specificity. We predict that these rearrangements occur stochastically in the parent population even in the absence of selective pressure. To characterize the occurrence of specific DNA rearrangements in unselected populations grown in succinate, a nested PCR approach was designed to target specific recurring junctions: JCN869, JCN950, JCN951 (as outlined in Figure 5.4). Nested PCR primers were designed to target a specific DNA rearrangement, and presence of a specific junction was determined in various cell populations. An example experiment, designed to detect JCN869, is shown in Figure 5.4. As expected, PCR products of the expected size were observed in the parent amplification mutant. PCR products were also apparent in DNA from unselected populations of wild-type ADP1 (Figure 5.4B) and in the ACN854 parent (data not shown). Samples spanning a range of DNA concentrations were used as template in separate PCR reactions. In the example experiment shown in Figure 5.4, template DNA from the amplification mutant, ACN869, yielded a PCR product when template DNA was diluted to 100 fg in the reaction. In contrast, PCR products were identified only in more concentrated DNA samples from the wild-type cells grown without selective pressure (20 ng to 200 ng). DNA sequencing of the cloned PCR products confirmed that the primers amplified a DNA fragment with a sequence identical to that of the junction identified in ACN869.

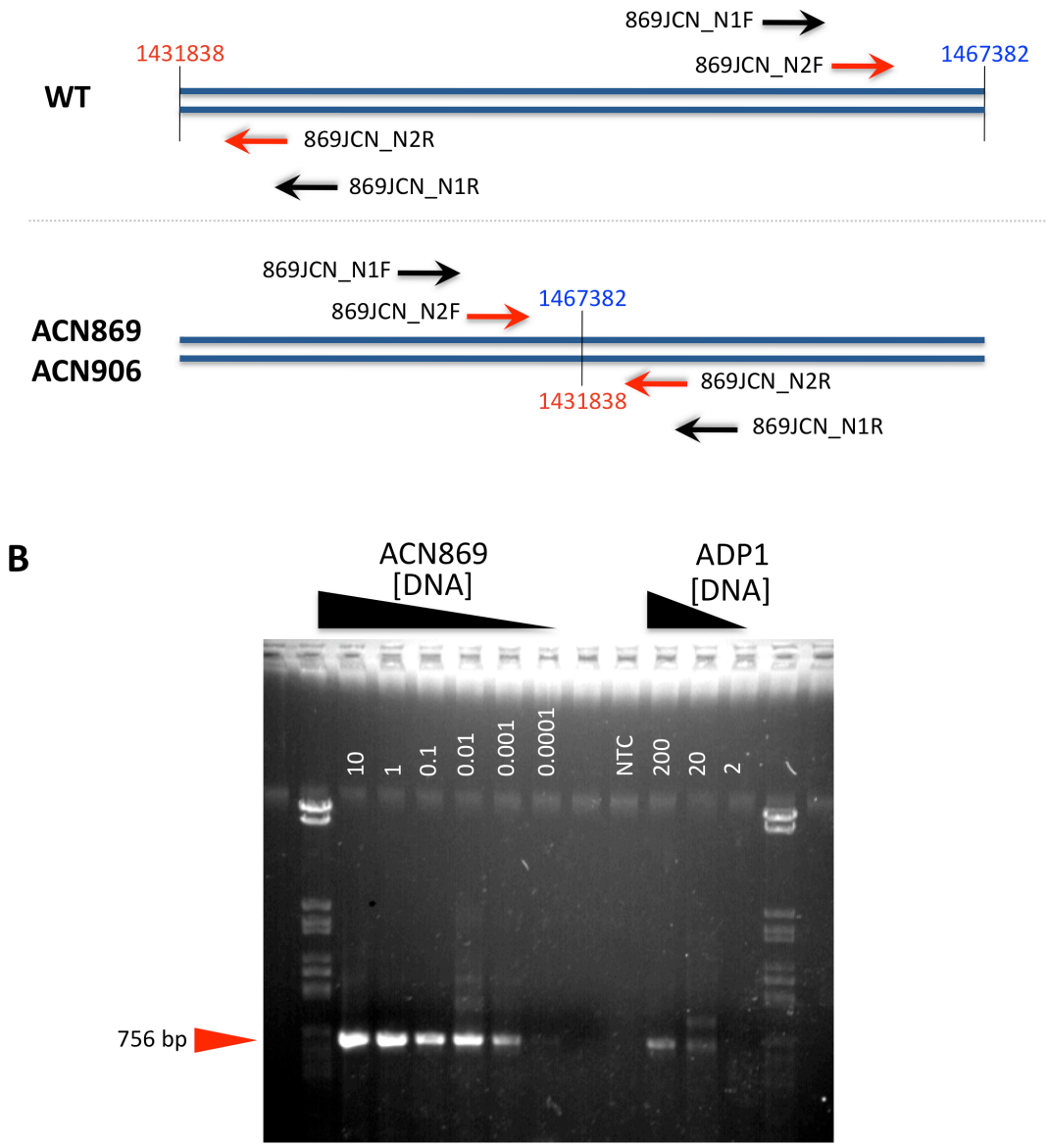


Figure 5.4: Nested PCR allows detection of specific DNA junctions in various cell populations. Panel A is a schematic representation of the nested PCR approach targeting a specific junction, JCN869. PCR primers should not amplify the wild-type chromosomal arrangement (top scenario), but will yield a product in strains carrying the specific tandem duplication (bottom scenario). In the first of two PCR reactions, DNA from the benzoate-grown amplification strain and DNA from succinate-grown populations of the wild-type strain is used as the template in reactions with the outer primers: 869JCN_N1F and 869JCN_N1R (black arrows). The resulting PCR product is used as the template in the second, nested reaction. Nested primers (red arrows) are designated 869JCN_N2F and 869JCN_N2R. In panel B, the results of one set of nested reactions are shown. DNA from the amplification strain, ACN869, was standardized and subjected to a series of tenfold dilutions (10 ng to 0.0001 ng, shown here). A product of the correct size is visible in concentrated (200 ng and 20 ng) samples of ADP1 DNA. NTC: No template control.

Analogous nested PCR experiments were designed to target JCN950 and JCN951. In both cases, PCR products of the expected size and sequence were detected in unselected populations of ADP1 and ACN854 (data not shown). In parallel experiments, each junction-specific primer set was used in a set of reactions with *E. coli* DH5α genomic DNA as the template. Junction primers did not generate a visible product in reactions with DNA from *E. coli*, thus serving as an important negative control. This result suggests that the specific genetic rearrangements are present in unselected *Acinetobacter* populations and are not an artifact of DNA cross-contamination during PCR experiments.

Use of anthranilate as a sole carbon source for the selection of amplification mutants

Growth on anthranilate provides an alternative strategy for selecting amplification mutants from the previously described ACN293 parent. Recall that ACN293 harbors a point mutation in the *benA* promoter, conferring high-level *ben* gene expression. Previous data (Reams and Neidle, 2004a) suggest that in the presence of benzoate, multiple copies of this promoter mutation are not tolerated and thus limit the site of duplication. This is likely due to the toxic accumulation of catechol due to an imbalance in *ben* and *cat* gene expression. However, this toxic metabolite accumulation would not occur in the absence of benzoate, the substrate of the *benABC* dioxygenase. To circumvent this limitation, anthranilate was used for carbon source selection using the rationale described below.

In ADP1, anthranilate is converted to catechol by a dioxygenase encoded by the *antABC* genes that are distant on the chromosome from the *ben-cat* cluster (Eby *et al*., 2001). Expression of *catA* and *catBCIJFD* are subsequently required to convert catechol

to tricarboxylic acid cycle intermediates for growth on anthranilate (See Figure 3.1 in Chapter 3). ACN293 does not grow on anthranilate, presumably because of insufficient expression of *catA* and *catBCIJFD* in the absence of BenM and CatM. The selection for growth on anthranilate (Ant^+) should be similar to that for benzoate, requiring increased expression of the *cat* genes through amplification.

To select mutants harboring chromosomal amplifications, independent cultures of ACN293 (~10^9 cells) were plated on solid medium with anthranilate as the sole carbon source. Spontaneous mutants able to grow on anthranilate (Ant^+) arose following incubation at 30°C for four to seven days. To assess the chromosomal configuration of mutants, Ant^+ derivatives were analyzed by PFGE. Approximately 20 independent mutants were characterized, and each harbored obvious rearrangements indicative of amplification (Figure 5.5B, for example). The amplicon endpoints (DNA junctions) were determined in several ACN293-derived Ant^+ mutants and are shown in Figure 5.5A. In each case DNA sequence identity at the recombination site was minimal, indicating the tandem duplication was likely formed via illegitimate recombination. Contrary to our prediction, the upstream DNA recombination site did not extend upstream of the *benA* promoter mutation in any characterized mutant, and thus did not expand the chromosomal region under study.

Modifications in the experimental system to investigate gene duplication and amplification throughout the A. baylyi *ADP1 chromosome*

Our experimental system has so far relied on examining rearrangements that surround the *ben* and *cat* genes in their normal chromosomal position

A

```
JCN871
CACCCGGGACAGCATTGTTGGTTAC TTTGTTTCCCAACATTAGTTTAAAT 1438285
TTGTTGGTATAGGTGTATTTAGTCA TTTGTTTCCCAACATTAGTTTAAAT JCN
TTGTTGGTATAGGTGTATTTAGTCA AACGTTGTAAAGGATACTGATTGAG 1708339

JCN872
AGAGTGGTCATGGTTTTCAACCTTA AATTTGGCGCTGCCGCTAATTTTGG 1438673
TTTTCAAATAATAAGTTGCAGCTTT AATTTGGCGCTGCCGCTAATTTTGG JCN
TTTTCAAATAATAAGTTGCAGCTTT TTTAAAGGCCCATTCTGGAATATCG 1632109

JCN873
GGCTCAAAAGTATGAAGCAAGCGAT TGCTGCGGAAAAAGAATTTGCATCG 1435533
ATGCAGCTGCTGACAACACGCCCTG TGCTGCGGAAAAAGAATTTGCATCG JCN
ATGCAGCTGCTGACAACACGCCCTG ATTGCACATGCTGGTGCCCAATAGT 1453815

JCN874
GGAATGGCGACGCCAGACGATCTT GA GGAGTTCCGTGCCTGTCAGGCTGG 1435325
GATAACTTAAGAGTTGCCAAGCAA GA GGAGTTCCGTGCCTGTCAGGCTGG JCN
GATAACTTAAGAGTTGCCAAGCAA GA TGCAACGGAGCAAGAGATGAAAGA 1511948

JCN875
AATGTTATGCACCTCAAGCCTCATT T TGGATGCCCGCGTGGGACGATAACG 1435706
TGTTCTTTCTAGCTGGCTATCCACA T TGGATGCCCGCGTGGGACGATAACG JCN
TGTTCTTTCTAGCTGGCTATCCACA T CTTGAAGAGTCATACCGTCTTGGTG 1537630

JCN877
TCTAAGGCGAATTTTGTTTTTTAAG A GATGACATATCACTCAGCTTAATTT 1439280
GTTTATACTGGCGATGACCTGTGGT A GATGACATATCACTCAGCTTAATTT JCN
GTTTATACTGGCGATGACCTGTGGT A TTTGTGCTGGATCGAATTATTATAA 1703630

JCN883
TTCCGCTATAAAAACAGTTACAGC TAT TTTGGCATGTCACGCTATGTAAT 1435966
TACTCTGGAACTGAATCAAAAAAT TAT TTTGGCATGTCACGCTATGTAAT JCN
TACTCTGGAACTGAATCAAAAAAT TAT AAAAAATGATCAGACAAGGGAAT 1709064
```

B

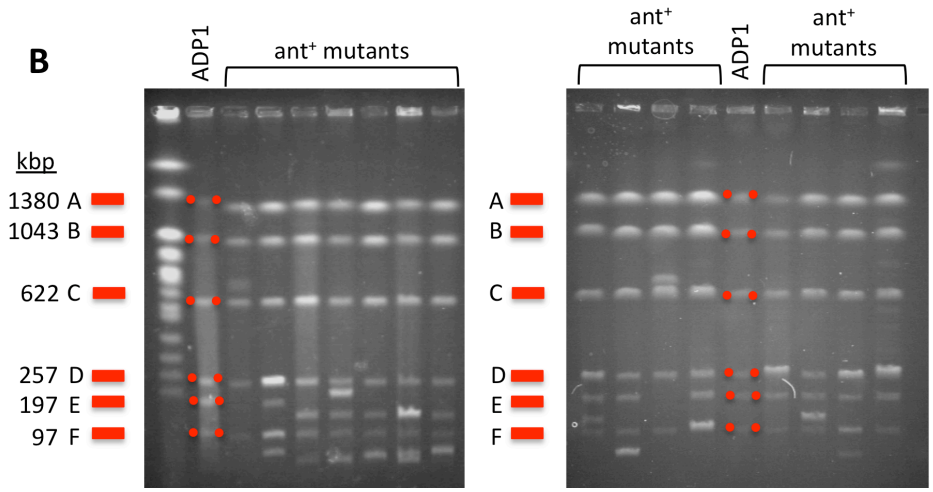


Figure 5.5: Anthranilate provides an alternative strategy for the selection of amplification mutants. Panel A aligns the DNA sequence involved in recombination to generate tandem duplication in Ant+ mutants derived from strain ACN293. For each alignment, the middle line (black text) indicates the novel DNA junction (JCN) from an amplification strain. The top line (red text) represents the parental upstream sequence, with the first displayed nucleotide identified by its position in the ADP1 genome sequence (GenBank entry: CR543861). The bottom line is the parental downstream sequence, with position indicated. Identical nucleotides at the DNA junction are represented by bold and underlined text. Homology in the vicinity of the recombination junction is represented by underlined text. Panel B reveals gross DNA rearrangements in amplification strains visualized by separation of NotI-digested genomic DNA via pulsed-field gel electrophoresis. Each lane represents an independently isolated mutant. For comparison, ADP1 DNA is digested with NotI, generating six fragments with sizes indicated (highlighted by red). Large chromosomal aberrations are seen in each mutant strain and are indicative of gene amplification. DNA was separated on a 1% gel made with low-melting point agarose and run with the following parameters: Voltage: 6 V, Switch time: 10-200 seconds, Run time: 24 hours.

(Reams & Neidle, 2004a). This method has focused on an area spanning less than 10% of the chromosome. To exploit the robust carbon source selection for further investigation of recombination events that occur elsewhere on the chromosome, a method was devised to relocate the *cat* genes to distant chromosomal positions. Irrespective of genomic location, in the absence of BenM and CatM, growth on benzoate should select multiple copies of *catA* and *catBCIJFD*. However, the DNA aligned during recombination to generate the duplication and subsequent amplification will vary with genomic position. This strategy allows a more comprehensive determination of the range and distribution of genetic amplification in ADP1.

To maintain the robust carbon source selection, the *catA* and *catBCIJFD* genes were relocated to different positions in the ADP1 chromosome. This involved stepwise genetic manipulation and was completed as follows: 1) the *cat* genes, *catA, catM* and *catBCIJFD*, were deleted from the native ADP1 chromosomal locus, 2) a plasmid was engineered to encode the full suite of *cat* genes flanked by unique restriction sites (pBAC841, Figure 5.6), 3) DNA from the target locus was PCR amplified and cloned into sites flanking the *cat* gene cassette, 4) the final plasmid was used as the donor in a transformation with an *Acinetobacter* recipient to introduce the *cat* gene cassette into the target locus via allelic exchange. A diagram outlining the engineering scheme is given in Figure 5.6 and details are provided in *Experimental Procedures*.

As described here, two different loci were targeted for insertion of the *cat* gene cassette, housed on plasmid pBAC841 (Figure 5.6). The first target locus (ACIAD0982, encoding the vanillate transporter *vanK*) is dispensable for growth on succinate and benzoate and is over 450 kbp removed from the native *ben-cat* locus (See Figure 5.7A).

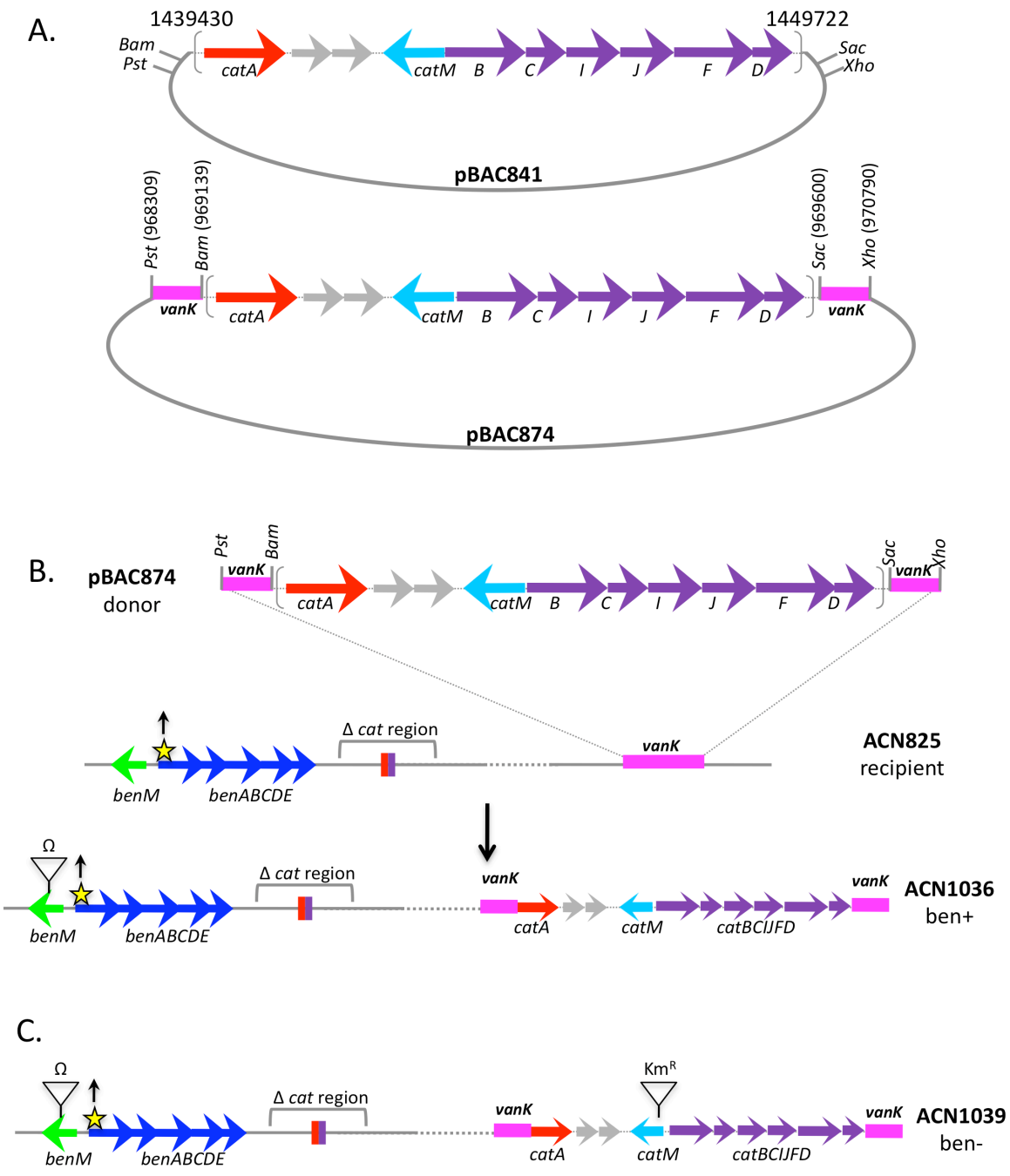

A.
1439430
1449722
Bam
Pst
catA
catM
B
C
I
J
F
D
Sac
Xho
pBAC841
Pst (968309)
Bam (969139)
Sac (969600)
Xho (970790)
vanK
catA
catM
B
C
I
J
F
D
vanK
pBAC874
B.
pBAC874
donor
Pst
Bam
vanK
catA
catM
B
C
I
J
F
D
Sac
vanK
Xho
Δ cat region
vanK
ACN825
recipient
benM
benABCDE
Ω
Δ cat region
vanK
vanK
ACN1036
ben+
benM
benABCDE
catA
catM
catBCIJFD
C.
Ω
Δ cat region
KmR
vanK
vanK
ACN1039
ben-
benM
benABCDE
catA
catM
catBCIJFD

Figure 5.6: Genetic engineering allows relocation of the *cat* gene selection. A. A universal plasmid construct (pBAC841) was engineered to encode the entire *cat* gene region (1439430-1449722 from GenBank entry CR543861) flanked by unique restriction sites: *Pst*I, *BamH*I, *Sac*I, and *Xho*I. DNA from the desired target locus (*vanK*, in this example) is PCR amplified and cloned into pBAC841 to generate the final shuttle plasmid. **B.** The final shuttle plasmid (pBAC874, in this example) is used as the donor DNA to insert the *cat* gene cassette into an *Acinetobacter* strain via allelic exchange. The recipient *Acinetobacter* strain, ACN825, harbors a full *cat* region deletion (Δ *cat*), a *benA* promoter mutation conferring high-level *benABCDE* expression (↑) and an inactivated *benM* (Ω). Correct integration of the *cat* gene cassette into the *vanK* locus of ACN825 generates a strain that grows on benzoate, ACN1036**. C.** Subsequent insertional inactivation of *catM* in ACN1036 generates the final strain, ACN1039. (A detailed description is provided in *Experimental Procedures*).

A plasmid, pBAC874, was engineered to carry the *cat* genes flanked by DNA from the *vanK* target locus. Using pBAC874, the entire *cat* gene region was inserted into *vanK* on the ACN825chromosome, generating strain ACN1036. Strain ACN1036 grows on benzoate indicating that the *catA*, *catM*, and *catBCIJFD* genes are functionally expressed in the new genetic context. Subsequent insertional inactivation of *catM* (*catM*::Km^R) in strain ACN1036 generates a benzoate-deficient derivative designated ACN1039. ACN1039 serves as the parent strain for selection of spontaneous mutants that acquire the ability to grow on benzoate.

Strain ACN1023 was similarly engineered to encode an intact *cat* gene region within the ACIAD3593 chromosomal locus (ACIAD indentifiers are derived from the genome annotation by Barbe *et al.*, 2004). The ACIAD3593 target locus, an 855 bp open reading frame encoding a predicted membrane protein, is ideal in that it is distant from the wild-type *ben-cat* cluster (Figure 5.7A) and dispensable when ADP1 is grown with succinate or benzoate as the sole carbon source (de Vallenet *et al*., 2008 and Craven, unpublished observation). ACN1023 grows on solid medium containing 1 mM benzoate as the sole carbon source indicating an intact and functional set of *cat* genes in the new chromosomal position. Subsequent insertional inactivation of *catM* (*catM*::Km^R) by allelic exchange precludes growth on benzoate and yields strain ACN1024. ACN1024 serves as the parent strain for isolation and characterization of spontaneous mutants that acquire the ability to grow on benzoate. A full description of the construction of ACN1024 is described in *Experimental Procedures*.

For selection of potential amplification mutants, individual colonies of ACN1039 or ACN1024 were grown to high cell density with succinate as the carbon source,

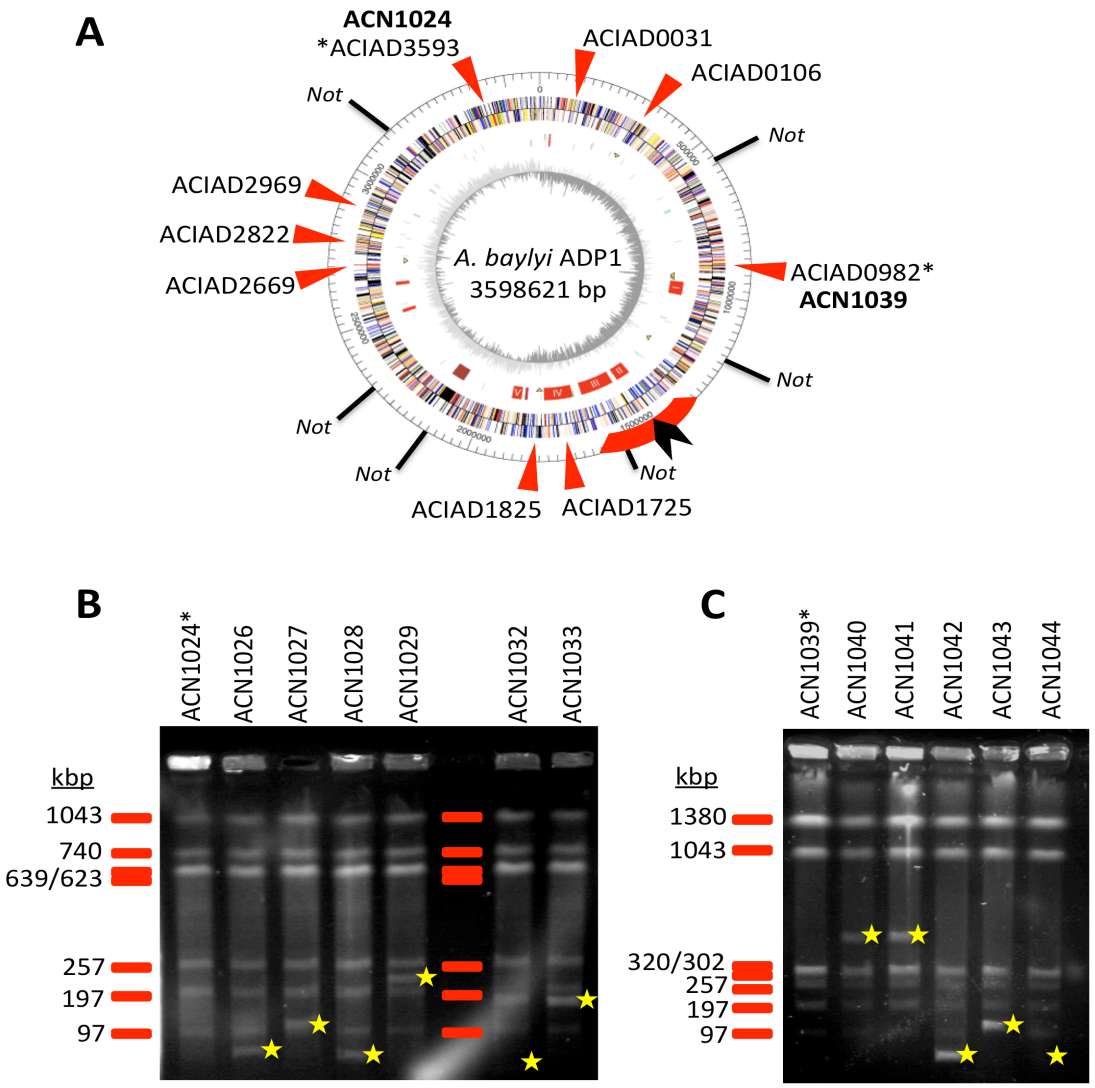


Figure 5.7: Relocation of the *cat* gene selection to varied chromosomal loci. In **A,** the ADP1 genome is shown, with loci currently being targeted for movement of the *cat* gene selection (red arrows). Engineering of two specific loci are complete and marked with an asterisk (*vanK* at open reading frame ACIAD0982 and a putative transporter at open reading frame ACIAD3593). The wild-type location of the *cat* genes is shown with the black arrowhead. *Not*I restriction sites (*Not*) are depicted for reference. **B**. PFGE analysis of the ACN1024 parent strain and benzoate-derived mutants. Genomic DNA from the ACN1024 parent generates seven fragments when digested with the *NotI* endonuclease (sizes are shown, left). In the benzoate^{+} mutants, large chromosomal aberrations (yellow stars) reveal amplification. **C.** PFGE analysis of the ACN1039 parent strain and benzoate-derived mutants. Genomic DNA from the ACN1039 parent generates seven fragments when digested with *NotI* (sizes are shown, left). Mutants derived from ACN1039 harbor apparent amplifications (yellow stars). In **B** and **C,** DNA was separated on a 1% gel made with low-melting point agarose and run with the following parameters: Voltage: 6 V, Switch time: 10-200 seconds, Run time: 24 hours.

concentrated, and spread (~5 × 10^9 cells) on solid benzoate medium. Small colonies arose after prolonged exposure to benzoate (5 to 14 days) and were streak purified to ensure a clonal population. In total, eight ACN1024-derived mutants were selected for further consideration (Table 5.1). Each was confirmed by phenotype: Sp^RSm^R, Km^R, Ben^+, $muconate^+$. PCR was used to confirm the *cat* region deletion. To visualize potential chromosomal rearrangements, total genomic DNA from Ben^+ mutants was digested with the *Not*I endonuclease and separated via PFGE. Results of the PFGE are shown in Figure 5.7B and confirm large-scale genomic rearrangements indicative of amplification. An analogous strategy was used to characterize $benzoate^+$ mutants derived from strain ACN1039. Analysis of five $benzoate^+$ mutants derived from ACN1039 reveal genomic rearrangements in these strains (Figure 5.7C). Taken together, these results suggest that amplification of the *cat* genes occurs at varied positions throughout the ADP1 genome. We propose to extend this strategy to addition loci as highlighted in Figure 5.7A, allowing a comprehensive survey of genetic duplication and amplification throughout the genome.

Discussion

Studies by Reams and Neidle suggest that tandem duplication of the *cat* region may occur by a position-specific illegitimate recombination mechanism. However, a point mutation in the parental strain artificially constrained the DNA region involved in recombination (Reams and Neidle, 2004a). This restriction may skew the nature of duplication sites under investigation. In the present study, various approaches were used to analyze duplication formation in the absence of constraint: alteration of carbon source

selection, utilization of a new parent strain for selection of amplification mutants, and movement of the *cat* genes to distant chromosomal loci.

Chromosomal amplifications lead to enhanced catabolism in Acinetobacter *strains*

Previously, benzoate was used as the selective pressure for isolation of mutants with chromosomal amplifications (Reams and Neidle, 2004a). The present study reveals that amplification of the *cat* genes can further allow growth on an alternative carbon source, anthranilate. Duplication formation in strains selected on anthranilate mimic those reported previously, showing little or no DNA sequence identity at the duplication junction (Figure 5.5). Interestingly, a large proportion of the ACN293-derived mutants characterized (4/7) amplified a very large chromosomal segment (265 kbp to 275 kbp) with a downstream amplicon terminus in the *pca* operon. A similar proportion of these events (45/104) was found in the expansive survey of benzoate mutants reported previously. These results are intriguing, as both the *pca* and *cat* operons are involved in aromatic compound catabolism and share regions of sequence homology. For example, *catIJF* and *pcaIJF* encode isozymes that are 98% identical in nucleotide sequence. Although no sequence similarity is evident at the exact duplication junction, the possibility exists that DNA interactions between regions of homology may serve as an anchor and facilitate a synapse between distant regions. Potential interactions between the *pca* and *cat* genes are currently being examined.

In a related set of experiments using the ACN854 parent, increased dosage of the *ben* and *cat* genes affords increased consumption of benzoate. Interestingly, despite the requirement for a larger amplicon in mutants derived from ACN854 than in ACN293, no detectable difference in selection frequency or growth rate was documented. More

importantly, the upstream duplication endpoints were in a chromosomal region not characterized in previous studies (Figure 5.8). Despite the absence of known constraints on amplicon endpoints, a surprising bias for illegitimate recombination persists. It is important to note that the apparent predisposition for recombination with the *pca* region, described above, is not apparent in mutants derived from ACN854. Rearrangements involving the distant *pca* region were not observed in any mutant. In fact, with one exception, amplicons in these strains were comparatively small, on average only 30 kbp (See Table 5.2). Different classes of recombination events may be favored under different selective pressures; however, a larger sample set of mutants will be required to form more concrete conclusions.

Duplication formation occurs preferentially at specific DNA sites

Amplification mutants were derived from the ACN854 parent strain and the precise DNA sequence at duplication join points determined (Figure 5.3). Interestingly, despite the small sample size (n=12 independent DNA junctions) a subset of duplications occurred precisely in multiple independent mutants. Qualitative nested PCR methods used to study three recurring duplications revealed that they occur stochastically in cell populations in the absence of selective pressure (Figure 5.4).

Site preference in duplication formation is documented in other organisms, often relying on recombination between sequence repeats (Romero and Palacios, 1997). Site preference in ADP1, however, appears unique. Obvious DNA sequence patterns associated with known site-specific recombinases or recombinational hotspots are not

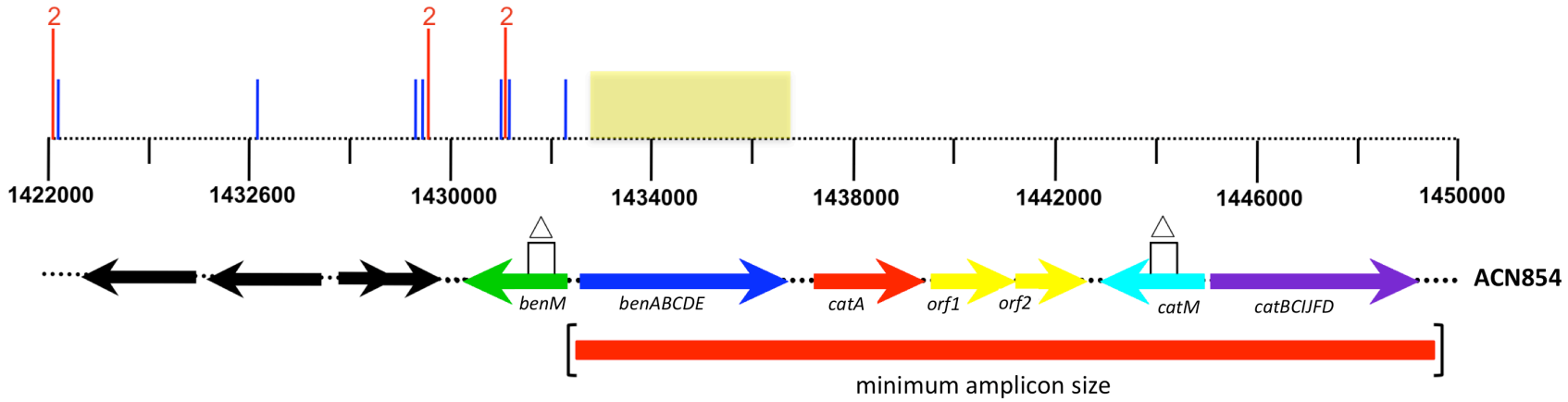


Figure 5.8: The distribution of DNA junctions in mutants derived from ACN854. The upstream recombination site is shown for each independently isolated amplification mutant. Blue vertical lines mark the position of the recombination site on the ADP1 genome (coordinates derived from GenBank entry CR543861). Red vertical lines indicate junctions that recur in independent isolates. The yellow box indicates the region in which all previous recombination events were found (Reams and Neidle, 2004). Note that downstream recombination sites are beyond the scope of this figure and varied from 5 to more than 70 kbp downstream of *catD*.

apparent. We are intrigued by the possibility that physical characteristics of the DNA, rather the base sequence *per se*, may influence the propensity for duplication formation at specific sites. For instance, a recent report of genetic duplication in *Plasmodium* revealed the presence of long monomeric A/T tracts at duplication endpoints (Nair *et al*., 2007). Intrinsic instability of AT-rich DNA tracts may simply make these regions more prone to DNA unwinding or double-strand break. If mechanisms of double-strand break and repair underlie genetic duplication in ADP1, DNA at immediate break points may be predicted to have lower local melting temperatures or other thermodynamic anomalies. To examine this prediction, local DNA melting temperature and inherent DNA curvature were analyzed here using the dan and banana programs, respectively, from the EMBOSS suite of bioinformatic tools (Rice *et al*., 2000). Junctions identified in more than one mutant (JCN869, JCN950, and JCN951) were analyzed. Average DNA melting temperature in the vicinity of DNA junction break points (using a 25 bp or 50 bp window centered around the junction point) did not vary significantly from that calculated for surrounding DNA sequence (data not shown). Nor were obvious anomalies in DNA curvature observed. Thus, the mechanism underlying PSIR remains unclear. More sophisticated bioinformatics approaches, combined with an increased sample size of DNA junction sequences may uncover underlying DNA patterns.

Amplification of the cat *genes can occur at varied loci throughout the ADP1 genome*

Here, a method was developed to examine gene amplification at positions throughout the ADP1 genomic landscape. As a proof of concept, strains ACN1024 and ACN1039 were engineered to examine amplification at the ACIAD3593 (genomic coordinate position 3 510 154) and ACIAD0982 (genomic coordinate position 968 383) loci, respectively. Mutants with

enhanced levels of benzoate consumption were isolated readily from the benzoate-deficient ACN1024 and ACN1039 parent strains. Analysis by PFGE (Figure 5.7B and 5.7C) confirms that the enhanced growth phenotype is correlated with large-scale chromosomal rearrangements indicative of amplification. Some conclusions can be drawn from PFGE analyses alone. First, amplification events are not identical as evidenced by the size variation seen in individual mutants (gold stars in Figure 5.7B and 5.7C). Secondly, the size of the amplicon can be estimated. Interestingly, the amplified region in some mutants is very large (>200 kbp in ACN1029; >400 kbp in ACN1040 and ACN1041). Amplicons in ACN1040 and ACN1041, therefore, represent the largest characterized in ADP1 to date. Future determination of the DNA junction sequences in these mutants will determine if illegitimate recombination remains a major source of genetic rearrangement throughout the genome.

The studies outlined here set the groundwork for a truly comprehensive evaluation of amplification events in ADP1. Constructs are currently underway that will enable relocation of the *cat* gene selection to ten additional loci. Examples of target loci are represented by red arrowheads in Figure 5.7A. We predict that genetic context, including proximity to the origin of replication or the presence of repetitive DNA, may alter the nature and frequency of genetic duplication at certain sites.

Thoughtful consideration of the target loci allows us to test related hypotheses. For example, the ACIAD3593 target locus was not only chosen because of its distance from the native *cat* locus, but also because of its proximity to the chromosomal *ampC* gene, encoding a functional β-lactamase. In ADP1, low-level *ampC* expression affords minimal resistance to ampicillin that can lead to enhanced resistance when overexpressed (Beceiro *et al*., 2007). In the present study, the *ampC* gene is within 5 kbp of the newly positioned *cat* genes in ACN1024.

Mutants of ACN1024 are selected on benzoate, and growth depends only on amplification of the *cat* genes. However, amplicons are large and may, by chance, include the adjacent *ampC*. We seek to test the hypothesis that coamplification of *ampC* will afford increased resistance to ampicillin in these mutants. The link between gene amplification and antibiotic resistance has been documented in diverse organisms including pathogenic *Acinetobacter baumannii* isolates. The tools developed in ADP1 may offer insight into events in other organisms.

Some genetic loci are prone to rearrangements, with duplication as frequent as 1 to 3 percent of all cells in a population (Anderson and Roth, 1981, Romero and Palacios, 1996). These recombination hotspots are almost exclusively defined by stretches of repetitive DNA sequence, long or short. The genomes of several genera including *Helicobacter*, *Mycoplasma*, *Rhizobium*, and *Streptomyces* maintain extensive repetitive DNA and thus are programmed for this type of plasticity (Aras *et al*., 2003). Extrapolating from patterns in these organisms, we postulate that some positions in the ADP1 genome may favor rearrangement. The ADP1 genome is marked by a small fraction, less than 2%, of repetitive DNA sequences. The longest repeat sequence is composed of two copies on an insertion sequence, IS1236, adjacent to the vanillate operon (ACIAD0982, See Figure 5.7A) and has previously been reported to underly extensive genetic scrambling (Barbe *et al.*, 2004). Relocation of the *cat* genes to the *vanK* locus in ACN1039 will, therefore, allow us to test multiple hypotheses: Does duplication of the chromosomal *van* locus occur at a high frequency? Are the duplications mediated by rearrangement between neighboring IS elements? Are there true recombinational hotspots within the ADP1 genome? Ongoing studies will address these questions.

The data reported here offer insight into genetic amplification in *Acinetobacter*, but also have the potential for extrapolation. In short, gene amplification is a widely conserved

mechanism and promotes genome diversification and adaptation in diverse organisms. The novel experimental tools developed here have expanded the ADP1 model system to allow a broad survey of genetic duplication and amplification throughout the chromosome and under various selective pressures. Inherent advantages of the ADP1 system overcome many of the difficulties that typically thwart investigation of this type genome plasticity.

Experimental Procedures

Bacterial strains and growth conditions

Acinetobacter strains used in this study are described in Table 5.1. *E. coli* strains DH5∝ and Top10 (Invitrogen) were used as plasmid hosts. Unless otherwise noted, *E. coli* was grown at 37°C in Luria-Bertani medium. *Acinetobacter* strains were cultured in minimal salts medium with succinate (10 mM), benzoate (1 mM), muconate (2 mM), or anthranilate (2 mM) as the carbon source and grown at 30°C (Shanley *et al.*, 2006). Antibiotics were added, as needed at the following final concentrations: kanamycin, 25 μg/ml; spectinomycin and streptomycin 15 μg/ml; ampicillin 150 μg/ml.

DNA manipulation and plasmid construction

Standard methods were used for molecular techniques including restricition digest, ligation, DNA purification, and *E. coli* transfomations (Sambrook *et al.*, 1989). DNA sequencing was performed using the University of Georgia Integrated Biotech Laboratories (IBL) core facility.

Engineering A. baylyi strain ACN854

Previously described strain ACN389 contains a 360 bp *benM* deletion within an otherwise wild-type background (Bundy *et al*, 2002). To engineer a *catM* deletion, ACN389 was used as the recipient in a cross with pBAC708, a plasmid harboring a *sacB*::Km^R marker within *catM*. Transformants in which the wild-type *catM* was replaced with the *sacB*::Km^R marker were selected on medium containing kanamycin and were sucrose sensitive (LB + 5% w/v sucrose) (Jones and Williams, 2003). The resulting strain was designated ACN843. Subsequently, the *catM*::*sacB* allele in ACN843 was replaced with the corresponding allele from pIGG6 (446 bp *catM* deletion, *catM*Δ5293) yielding strain ACN854. The presence of *ΔbenM* and *ΔcatM* alleles in ACN854 was checked by phenotype (benzoate$^-$, muconate$^-$) and confirmed by PCR (data not shown). Subsequently, strain ACN854 was used as the parent strain for isolation of mutants with genomic amplifications of the *ben-cat* region.

PCR assay investigates the occurrence of specific duplication junctions

A nested PCR method was used to detect the spontaneous occurrence of a specific tandem duplication in various cell populations (Figure 5.4A). Nested PCR primers were designed to target a specific recurring junction: JCN869, JCN950, or JCN951. Oligonucleotide primers were synthesized by Integrated DNA Technologies (IDT, Coralville, IA) and sequences of each are shown in below:

869JCN_N1F: GCAGAACTTCCACAGGCCTTACGT

869JCN_N1R: GCCGACACGTCCCATGC

869JCN_N2F: GGGAACACACGGCGG

869JCN_N2R: GGCTGATACTGATACGACGG

JCN950_N1F: CAAGAGGCGACGCTGATG

JCN950_N1R: GATGGCTAGTATTAATGACGGGAATC

JCN950_N2F: GCCCTGACTTCTGTAGGTCTTGG

JCN950_N2R: GCACAGCCGCCCTTAAG

951JCN_N1F: GCCCGCGGTCTTACAA

951JCN_N1R: CAATATAGGGCCGTTT

951JCN_N2F: GAGTATGCCCGGTAGATG

951JCN_N2R: GATTAAACGCATCCCAT

To detect specific DNA junctions in various cell populations, genomic DNA was isolated from benzoate-grown amplification strains and unselected succinate-grown populations of ADP1 and ACN854 using the Easy DNA (Invitrogen) purification protocol. For the first PCR in the set of two nested reactions, a range of genomic DNA concentrations (100 fg to 200 ng) was used as the template in reactions with the outermost primer set. Reaction products were purified using the QIAquick PCR purification kit protocol (Qiagen). A 2-µl aliquot from the first PCR was used as the template for the nested reaction. Final PCR products were visualized on 0.7% agarose stained with ethidium bromide. To confirm that correctly-sized bands were identical in sequence to the targeted junction, PCR products were cloned into the pCR2.1-TOPO vector using the TOPO TA Cloning Kit (Invitrogen). Plasmids were sequenced using the T7 promoter universal priming site located on the vector.

Creation of cat gene shuttle plasmid

A plasmid, designated pBAC841, was engineered to carry the entire wild-type *cat* gene region from the ADP1 chromosome (genomic coordinates: 1439435-1449713) flanked by unique

restriction sites to facilitate cloning and insertion of the cassette at varied chromosomal loci. First, an ~2.2-kbp *Eco*RV to *Bgl*II fragment containing the *catA* gene and its upstream promoter region was cloned into plasmid, pBAC184. The resulting plasmid was designated pBAC832. To add additional restriction sites, an ~500-bp *Dra*III to *Eco*RI fragment from pET21b (Invitrogen) was ligated into pBAC832, generating pBAC840. Finally, pBAC840 was linearized with *Cla*I and transformed into wild-type ADP1 to allow capture of the remaining *cat* region DNA via the gap-repair method (Gregg-Jolly and Ornston, 1990). The resulting plasmid, pBAC841, contains *catA*, ACIAD1443-1444, *catM*, and the *catBCIJFD* operon flanked by unique restriction sites (*Pst*I and *Bam*HI flanking *catA* and *Sac*I and *Xho*I flanking *catD*). For a detailed representation of pBAC841, please see Figure 5.6A.

pBAC841 served as the universal starting plasmid for the movement of the *cat* gene region to each chromosomal target locus. Here, details of the scheme used to engineer pBAC874, a construct used to insert the *cat* gene region into the chromosomal *vanK* locus, will be outlined. Other shuttle constructs were engineered analogously. Briefly, a PCR primer pair was designed to amplify an ~1 kbp product within the *vanK* target gene, while introducing *Pst*I and *Bam*HI restriction sites at the 5' and 3' ends, respectively. The PCR product was digested with *Pst*I and *Bam*HI and ligated into similarly digested pBAC841, yielding plasmid pBAC873. A second *vanK* PCR product was generated with primers designed to introduce *Sac*I and *Xho*I restriction sites at the 5' and 3' ends, respectively. This PCR product was digested with *Sac*I and *Xho*I and ligated into similarly digested pBAC873. The resulting plasmid was designated pBAC874 and has the following configuration: *vanK*:*catA*-ACIAD1443-1444-*catM*-*catBCIJFD:vanK* (See Figure 5.6A). A plasmid construct designed to move the *cat* region to another locus, an open

reading frame designated ACIAD3593 (pBAC850), was similarly engineered and is discussed in the text.

Movement of the cat gene cassette to varied positions throughout the ADP1 genome

The *catA-catD* genes were first deleted from the native ADP1 chromosomal locus. Briefly, a plasmid-borne deletion encompassing the *cat* genes (genomic coordinates: 1439923-1449545) was engineered via overlap-extension PCR. The resulting plasmid, pBAC761, was linearized with *Nde*I and used to transform ACN814, a strain with a *sacB*::Km^R counter-selectable marker in the chromosomal *catM* gene. Transformants in which the marker was replaced with the *cat* region deletion were selected on LB medium containing 5% sucrose, and presence of the deletion was confirmed by PCR (data not shown). The resulting strain, ACN825, maintains the *benM* insertion (ΩSpSm) and the *benA* promoter mutation (allele *benA*5147) from ACN147 and contains a deletion of the region encompassing *catA*, ACIAD1443-1444, *catM*, and the *catBCIJFD* operon.

ACN825 served as the parent for construction of strains in which the full wild-type *cat* region was inserted into distant chromosomal loci. As above, construction of strains ACN1036 and ACN1038, which harbor the *cat* genes within the chromosomal *vanK* locus, are described in detail. Analogous strains in which the *cat* region is shuttled to ACIAD3593 were engineered similarly. First, pBAC874 (above, and Figure 5.6) was linearized with *PstI* and used as the donor in a transformation with ACN825. Transformants in which the full wild-type *cat* region was inserted into the chromosomal *vanK* locus were selected by the ability of cells to form colonies on solid medium with benzoate (1 mM) as the sole carbon source. Correct integration of the *cat* genes was confirmed by PCR and PFGE.

Next, the CatM transcriptional regulator was inactivated by insertion of a *catM*::Km^R cassette into strain ACN1036. Briefly, pBAC6A, a plasmid harboring *catM*::Km^R, was linearized with *Xho*I and used as the donor in a transformation with ACN1036. Transformants in which the plasmid-borne *catM*::Km^R allele replaced the chromosomal *catM* gene of ACN1036 were selected on medium containing kanamycin (25 μg/L). The resulting strain, designated ACN1039 (Figure 5.6C), does not grow on benzoate and serves as the parent strain for selection of spontaneous benzoate$^+$ amplification mutants

Quantitative real-time PCR allows determination of amplicon copy number

A TaqMan 5' exonuclease assay was used to determine amplicon copy number in mutant strains harboring genomic amplifications. In these assays, the relative abundance of *catA*, was compared to that of single-copy endogenous control gene, *antA*. The primer-probe sets for detection of *catA* and *antA* were designed using the PrimerExpress software package (Applied Biosystems). Primers used for detection of *catA* and *antA* are listed below:

catA_forward: CATTACCTCGATATGCGTATGGAT

catA_reverse : CGTGGTGTCGCATTTTCAAT

antA_forward: TGCCGTACTTGGTGGAAAAGT

antA_reverse: CGGTGCATGGCCCATT

The TaqMan probes for detection of *catA* and *antA* are labeled at the 5' end with fluorescent reporters, FAM (6-carboxyfluorescein) and VIC, respectively. Total DNA from wild-type ADP1 and amplification mutant strains was purified using the Easy DNA purification protocol as specified by the manufacturer (Invitrogen). For quantification of the *catA* target, a standard curve was generated using ADP1 genomic DNA as the template for PCR (1000 ng to 0.01 ng). Using a

Relative Standard Curve method, the amount of the *catA* target is determined relative to a single-copy endogenous control (*antA*). For ADP1, the *catA*:*antA* ratio is assumed to be 1:1. Ratios of *catA*:*antA* in amplification mutants are, therefore, equivalent to the amplicon copy number in each mutant.

A transformation assay exploits the natural competence of ADP1 and facilitates determination of exact duplication sites

The DNA sequence of chromosomal duplication sites (junction DNA) was determined, as previously described, with some modifications (Reams and Neidle, 2003). Briefly, genomic DNA was isolated from amplification mutants grown on benzoate (1 mM), muconate (2 mM), or anthranilate (1 mM). Plasmid libraries were generated using the pZErO-2 cloning vector (Invitrogen) digested with *Eco*RI, *Hind*III, or *Asp*718 and similarly digested chromosomal DNA from independent mutants. Plasmid libraries were housed in *E. coli* Top10 (Invitrogen). Plasmid libraries were screened for DNA capable of conferring growth on the appropriate carbon source: benzoate, muconate, or anthranilate. Briefly, the benzoate-deficient parent strain (ACN293, ACN854, ACN1024) is spread on solid minimal medium supplemented with succinate. Immediately, individual *E. coli* colonies from the pZErO library are transferred to (1) an LB-kanamycin master plate and then (2) the succinate medium atop the previously spread *Acinetobacter* parent. *E. coli* does not grow well on succinate medium and will donate plasmid DNA to transform adjacent *Acinetobacter* cells. Following overnight incubation at 37°C, *Acinetobacter* cells are transferred from succinate to selective medium (benzoate, muconate, or anthranilate). Recombinant plasmids harboring junction DNA will confer growth on the

selective medium and are subsequently purified from the *E. coli* housed on a master plate. DNA sequencing allows determination of the exact DNA sequence at the site of recombination.

CHAPTER 6

CONCLUSIONS

The work presented in this dissertation highlights details of *Acinetobacter* regulation and physiology. In addition, the topics explored here have broader implications in prokaryotic physiology and may impact biotechnology and medicine. Some future implications and a brief summary are provided here.

Unraveling details of the LysR family of transcriptional regulators

Despite pervasiveness in diverse bacteria, many details of LTTR structure and function remain unknown. Structural studies have long been obscured by low solubility. Thus, atomic-level structures exist for very few representatives and generally include only the effector-binding domain: BenM and CatM, CysB, Cbl, and OxyR, and others (reviewed in Maddocks *et al*., 2008). Until recently, the complete structure of only one LTTR had been extensively characterized, that of CbnR for *Ralstonia eutropha* (Muraoka *et al*., 2003). However, in a recent report by Sainsbury and colleagues, a full-length structure of the CrgA regulator from *Neisseria meningitidis* is presented (Sainsbury *et al*., 2009). CrgA modulates expression of the adjacent *mdaB* gene, encoding a putative quinone NADPH oxidoreductase, and may be involved in oxidative stress response. The structure of CrgA revealed the assembly of the protein into an octameric ring, in marked contrast to the tetrameric form assumed to be the physiological oligomer for most LTTRs. Because structural data for this class are so sparse, it is intriguing to postulate that CrgA is a member of a distinct subclass of regulators that function as octamers.

Few examples address the relationship between LTTR structure and function. Even fewer address structural determinants of effector binding. With regard to the BenM-CatM regulon of ADP1, collaboration affords the rare combination of extensive *in vivo* and structural data (Ezezika *et al*., 2007a; Clark *et al*., 2004a; Bundy *et al*., 2002). Building on these data, recent work reveals the full-length structures of BenM and of the variant BenM(E226K) described in Chapter 3 (Ruangprasert and Momany, unpublished data). Additional structures of BenM and CatM are underway with cognate promoter DNA (Alanazi and Momany, unpublished data). In short, the BenM-CatM regulon of ADP1 affords creative experimental opportunities that continue to elucidate basic details of the diverse LysR family of regulators.

While expanding the general understanding of the LysR family, approaches outlined in Chapter 3 of this dissertation have further implications in biotechnology. Bacterial regulatory cascades are appreciated for their potential in diverse applications including synthetic biology, biodegradation, and the development of biosensors (reviewed in Galvao and de Lorenzo, 2006; Tropel and van der Meer, 2004). Here, predictions from structural data were used to successfully alter the effector specificity of the BenM regulator *in vivo* (Ezezika *et al.* 2007a; Craven *et al*., 2009). Less-directed approaches including error-prone PCR, simple selection, and computational methods have shown moderate success in engineering proteins with enhanced functions or novel specificities (Galvao and Lorenzo, 2006). However, rational engineering is vastly improved with structural data, as evidenced in Chapter 3. Extension of similar methods may be useful in metabolic applications that yield specific regulators with higher sensitivities or yield variant regulators with specificity for toxic or even xenobiotic compounds.

Broader implications for genetic amplification in Acinetobacter *and other bacteria*

Genetic amplification often gives rise to clinically relevant consequences (reviewed in Craven & Neidle, 2007), and has recently been suggested to underlie increased antibiotic resistance in *Acinetobacter* species (Beceiro *et al*., 2007; Bertini *et al.*, 2007). *Acinetobacter baumannii* strains, for example, are important opportunistic pathogens and are often resistant to a wide range of antibiotics including β-lactams, aminoglycosides, fluoroquinilones, and carbapenems (Navon-Venezia *et al.*, 2005; Bertini *et al.*, 2007). Although acinetobacters may develop resistance through a variety of mechanisms, gene dosage effects associated with amplification of chromosomal or plasmid-borne resistance determinants require closer attention. For example, the $bla_{\text{OXA-58}}$ gene, encoding a carbapenem-hydrolyzing oxacillinase, occurs naturally in many *A. baumannii* species. These genes are often plasmid-borne and are usually weakly expressed. Clinically resistant strains isolated from a University Hospital in Rome, Italy, revealed strains harboring multiple copies of $bla_{\text{OXA-58}}$. Increased copy number in these isolates depended on duplication between neighboring IS elements, and significantly increased resistance in clinical strains (Bertini *et al.*, 2007).

Reports of gene amplifications and the emergence of antibiotic resistance inundate the current literature. Some examples alter the structure or availability of the drug directly, including amplification of β-lactamases or drug transporters (Sun *et al.*, 2009; Nicoloff *et al*., 2006). Other amplifications afford resistance through increased levels of the antibiotic target or by bypassing the targeted cellular pathway entirely (Nilsson *et al*., 2006; Brochet *et al.*, 2008). In any case, the high steady-state frequency of duplications (10^{-2} to 10^{-5} depending on genomic location) makes amplification much more likely in a population than simple point mutation. Adaptation through gene dosage is not restricted to antibiotic resistance, but instead is applicable to any case in

which increased activity of some gene provides a fitness advantage. Gene amplification, therefore, represents a relatively underappreciated mechanism of bacterial adaptation. The genetic tools developed to study these phenomena in ADP1 are of broad importance.

REFERENCES

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

Abele-Horn, M., U. Vogel, I. Klare, C. Konstabel, R. Trabold, R. Kurihara, W. Witte, W. Kreth, P. G. Schlegel & H. Claus, (2006) Molecular epidemiology of hospital-acquired vancomycin-resistant enterococci. *J. Clin. Microbiol.* **44**: 4009-4013.

Akakura, R. & S. C. Winans, (2002a) Constitutive mutations of the OccR regulatory protein affect DNA bending in response to metabolites released from plant tumors. *J Biol Chem* **277**: 5866-5874.

Akakura, R. & S. C. Winans, (2002b) Mutations in the *occQ* operator that decrease OccR-induced DNA bending do not cause constitutive promoter activity. *J Biol Chem* **277**: 15773-15780.

Albertson, D. G., (2006) Gene amplification in cancer. *Trends Genet* **22**: 447-455.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers & D. J. Lipman, (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.

Anderson, P. & J. Roth, (1981) Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (rrn) cistrons. *Proc. Natl. Acad. Sci. USA* **78**: 3113-3117.

Andersson, D. I., (2006) The biological cost of mutational antibiotic resistance: any practical conclusions? *Current Opinion In Microbiology* **9**: 461-465.

Apweiler, R., A. Bairoch & C. H. Wu, (2004) Protein sequence databases. *Curr. Opin. Chem. Biol.* **8**: 76-80.

Aras, R. A., W. Fischer, G. I. Perez-Perez, M. Crosatti, T. Ando, R. Haas & M. J. Blaser, (2003a) Plasticity of repetitive DNA sequences within a bacterial (Type IV) secretion system component. *J. Exp. Med.* **198**: 1349-1360.

Aras, R. A., J. Kang, A. I. Tschumi, Y. Harasaki & M. J. Blaser, (2003b) Extensive repetitive DNA facilitates prokaryotic genome plasticity. *Proc Natl Acad Sci U S A* **100**: 13579-13584.

Aravind, L., V. Anantharaman, S. Balaji, M. M. Babu & L. M. Iyer, (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**: 231-262.

Barbe, V., D. Vallenet, N. Fonknechten, A. Kreimeyer, S. Oztas, L. Labarre, S. Cruveiller, C. Robert, S. Duprat, P. Wincker, L. N. Ornston, J. Weissenbach, P. Marliere, G. N. Cohen & C. Medigue, (2004) Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Res.* **32**: 5766-5779.

Bokal, A. J., W. Ross, T. Gaal, R. C. Johnson & R. L. Gourse, (1997) Molecular anatomy of a transcription activation patch: FIS-RNA polymerase interactions at the *Escherichia coli rrnB* P1 promoter. *Embo J.* **16**: 154-162.

Bott, M., M. Meyer & P. Dimroth, (1995) Regulation of anaerobic citrate metabolism in *Klebsiella pneumoniae*. *Mol. Microbiol.* **18**: 533-546.

Bowater, R. & A. J. Doherty, (2006) Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet* **2**: e8.

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

Brzostowicz, P. C., A. B. Reams, T. J. Clark & E. L. Neidle, (2003) Transcriptional cross-regulation of the catechol and protocatechuate branches of the beta-ketoadipate pathway contributes to carbon source-dependent expression of the *Acinetobacter* sp. strain ADP1 *pobA* gene. *Appl. Environ. Microbiol.* **69**: 1598-1606.

Buck, M. & M. K. Rosen, (2001) Structural biology. Flipping a switch. *Science* **291**: 2329-2330.

Bundy, B. M., A. L. Campbell & E. L. Neidle, (1998) Similarities between the *antABC*-encoded anthranilate dioxygenase and the *benABC*-encoded benzoate dioxygenase of *Acinetobacter* sp. strain ADP1. *J. Bacteriol.* **180**: 4466-4474.

Bundy, B. M., L. S. Collier, T. R. Hoover & E. L. Neidle, (2002) Synergistic transcriptional activation by one regulatory protein in response to two metabolites. *Proc Natl Acad Sci U S A* **99**: 7693-7698.

Bykowski, T., J. R. van der Ploeg, R. Iwanicka-Nowicka & M. M. Hryniewicz, (2002) The switch from inorganic to organic sulphur assimilation in *Escherichia coli*: adenosine 5'-phosphosulphate (APS) as a signalling molecule for sulphate excess. *Mol. Microbiol.* **43**: 1347-1358.

Cebolla, A., C. Sousa & V. de Lorenzo, (1997) Effector specificity mutants of the transcriptional activator NahR of naphthalene degrading *Pseudomonas* define protein sites involved in binding of aromatic inducers. *J Biol Chem* **272**: 3986-3992.

Cebolla, A., C. Sousa & V. de Lorenzo, (2001) Rational design of a bacterial transcriptional cascade for amplifying gene expression capacity. *Nucleic Acids Res* **29**: 759-766.

Cerquetti, M., R. Cardines, M. L. Ciofi Degli Atti, M. Giufre, A. Bella, T. Sofia, P. Mastrantonio & M. Slack, (2005) Presence of multiple copies of the capsulation b locus in invasive

Haemophilus influenzae type b (Hib) strains isolated from children with Hib conjugate vaccine failure. *J. Infect. Dis* **192**: 819-823.

Cerquetti, M., R. Cardines, M. Giufre, A. Castella, M. Rebora, P. Mastrantonio & M. L. Ciofi Degli Atti, (2006a) Detection of six copies of the capsulation b locus in a *Haemophilus influenzae* type b strain isolated from a splenectomized patient with fulminant septic shock. *J. Clin. Microbiol.* **44**: 640-642.

Cerquetti, M., R. Cardines, M. Giufre, T. Sofia, F. D'Ambrosio, P. Mastrantonio & M. L. degli Atti, (2006b) Genetic diversity of invasive strains of *Haemophilus influenzae* type b before and after introduction of the conjugate vaccine in Italy. *Clin. Infect. Dis.* **43**: 317-319.

Chang, C. H., Y. C. Chang, A. Underwood, C. S. Chiou & C. Y. Kao, (2007) VNTRDB: a bacterial variable number tandem repeat locus database. *Nuc. Acids Res.* **35**: D416-421.

Chen, X. C., J. Feng, B. H. Hou, F. Q. Li, Q. Li & G. F. Hong, (2005) Modulating DNA bending affects NodD-mediated transcriptional control in *Rhizobium leguminosarum*. *Nucleic Acids Res* **33**: 2540-2548.

Chohnan, S., Y. Kurusu, H. Nishihara & Y. Takamura, (1999) Cloning and characterization of *mdc* genes encoding malonate decarboxylase from *Pseudomonas putida*. *FEMS Microbiol. Lett.* **174**: 311-319.

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

Clark, T., S. Haddad, E. Neidle & C. Momany, (2004a) Crystallization of the effector-binding domains of BenM and CatM, LysR-type transcriptional regulators from *Acinetobacter* sp. ADP1. *Acta Crystallogr D Biol Crystallogr* **60**: 105-108.

Clark, T. J., C. Momany & 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.

Clark, T. J., R. S. Phillips, B. M. Bundy, C. Momany & E. L. Neidle, (2004b) Benzoate decreases the binding of *cis,cis*-muconate to the BenM regulator despite the synergistic effect of both compounds on transcriptional activation. *J. Bacteriol.* **186**: 1200-1204.

Clewell, D. B., (1990) Movable genetic elements and antibiotic resistance in enterococci. *Eur. J. Clin. Microbiol. Infect. Dis.* **9**: 90-102.

Clough, S. J., K. E. Lee, M. A. Schell & T. P. Denny, (1997) A two-component system in *Ralstonia (Pseudomonas) solanacearum* modulates production of PhcA-regulated virulence factors in response to 3-hydroxypalmitic acid methyl ester. *J Bacteriol* **179**: 3639-3648.

Collier, L. S., G. L. Gaines, 3rd & E. L. Neidle, (1998) Regulation of benzoate degradation in *Acinetobacter* sp. strain ADP1 by BenM, a LysR-type transcriptional activator. *J Bacteriol* **180**: 2493-2501.

Collier, L. S., N. N. Nichols & E. L. Neidle, (1997) *benK* encodes a hydrophobic permease-like protein involved in benzoate degradation by *Acinetobacter* sp. strain ADP1. *J. Bacteriol.* **179**: 5943-5946.

Collins, N. E., J. Liebenberg, E. P. de Villiers, K. A. Brayton, E. Louw, A. Pretorius, F. E. Faber, H. van Heerden, A. Josemans, M. van Kleef, H. C. Steyn, M. F. van Strijp, E. Zweygarth, F. Jongejan, J. C. Maillard, D. Berthier, M. Botha, F. Joubert, C. H. Corton, N. R. Thomson, M. T. Allsopp & B. A. Allsopp, (2005) The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proc. Natl. Acad. Sci. U.S.A* **102**: 838-843.

Cosper, N. J., L. S. Collier, T. J. Clark, R. A. Scott & E. L. Neidle, (2000) Mutations in *catB*, the gene encoding muconate cycloisomerase, activate transcription of the distal *ben* genes and contribute to a complex regulatory circuit in *Acinetobacter* sp. strain ADP1. *J Bacteriol* **182**: 7044-7052.

Craven, S. H., O. C. Ezezika, C. Momany & E. L. Neidle, (2008) LysR homologs in Acinetobacter: Insights into a diverse and prevalent family of transcriptional regulators. In: *Acinetobacter* Molecular Biology. U. Gerischer (ed). Norfolk UK: Caister Academic Press, pp. 163-202.

Craven, S. H. & E. L. Neidle, (2007) Double trouble: medical implications of genetic duplication and amplification in bacteria. *Future Microbiol* **2**: 309-321.

D'Argenio, D. A., A. Segura, W. M. Coco, P. V. Bunz & L. N. Ornston, (1999) The physiological contribution of *Acinetobacter* PcaK, a transport system that acts upon protocatechuate, can be masked by the overlapping specificity of VanK. *J. Bacteriol.* **181**: 3505-3515.

Dal, S., I. Steiner & U. Gerischer, (2002) Multiple operons connected with catabolism of aromatic compounds in Acinetobacter sp. strain ADP1 are under carbon catabolite repression. *J Mol Microbiol Biotechnol* **4**: 389-404.

Dal, S., G. Trautwein & U. Gerischer, (2005) Transcriptional organization of genes for protocatechuate and quinate degradation from *Acinetobacter* sp. strain ADP1. *Appl. Environ. Microbiol.* **71**: 1025-1034.

Dangel, A. W., J. L. Gibson, A. P. Janssen & F. R. Tabita, (2005) Residues that influence in vivo and in vitro CbbR function in *Rhodobacter sphaeroides* and identification of a specific region critical for co-inducer recognition. *Mol Microbiol* **57**: 1397-1414.

Daunert, S., G. Barrett, J. S. Feliciano, R. S. Shetty, S. Shrestha & W. Smith-Spencer, (2000) Genetically engineered whole-cell sensing systems: coupling biological recognition with reporter genes. *Chem Rev* **100**: 2705-2738.

de Bruin, O. M., J. S. Ludu & F. E. Nano, (2007) The Francisella pathogenicity island protein IglA localizes to the bacterial cytoplasm and is needed for intracellular growth. *BMC Microbiol.* **7**: 1.

de Castro, L. A., T. Rodrigues Pedroso, S. S. Kuchiishi, M. Ramenzoni, J. D. Kich, A. Zaha, M. Henning Vainstein & H. Bunselmeyer Ferreira, (2006) Variable number of tandem aminoacid repeats in adhesion-related CDS products in *Mycoplasma hyopneumoniae* strains. *Vet. Microbiol.* **116**: 258-269.

Dean, C. R., S. Narayan, D. M. Daigle, J. L. Dzink-Fox, X. Puyang, K. R. Bracken, K. E. Dean, B. Weidmann, Z. Yuan, R. Jain & N. S. Ryder, (2005) Role of the AcrAB-TolC efflux pump in determining susceptibility of *Haemophilus influenzae* to the novel peptide deformylase inhibitor LBM415. *Antimicrob. Agents Chemother.* **49**: 3129-3135.

Dean, C. R., S. Narayan, J. Richards, D. M. Daigle, S. Esterow, J. A. Leeds, H. Kamp, X. Puyang, B. Wiedmann, D. Mueller, H. Voshol, J. van Oostrum, D. Wall, J. Koehn, J. Dzink-Fox & N. S. Ryder, (2007) Reduced susceptibility of *Haemophilus influenzae* to the peptide deformylase inhibitor LBM415 can result from target protein overexpression due to amplified chromosomal *def* gene copy number. *Antimicrob. Agents Chemother.* **51**: 1004-1010.

DeLano, W. L., (2002) *The Pymol User's Manual*. DeLano Scientific, San Carlos, CA.

Diaz, E., A. Ferrandez & J. L. Garcia, (1998) Characterization of the *hca* cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid in *Escherichia coli* K-12. *J. Bacteriol.* **180**: 2915-2923.

Dmitrova, M., G. Younes-Cauet, P. Oertel-Buchheit, D. Porte, M. Schnarr & M. Granger-Schnarr, (1998) A new LexA-based genetic system for monitoring and analyzing protein heterodimerization in *Escherichia coli*. *Mol. Gen. Genet.* **257**: 205-212.

Dove, S. L. & A. Hochschild, (2005) How transcription can be regulated in bacteria. In: The Bacterial Chromosome. N. P. Higgins (ed). Washington, DC: ASM Press, pp. 297-310.

Edlund, T. & S. Normark, (1981) Recombination between short DNA homologies causes tandem duplication. *Nature* **292**: 269-271.

Eggeling, L. & H. Sahm, (2003) New ubiquitous translocators: amino acid export by *Corynebacterium glutamicum* and *Escherichia coli*. *Arch. Microbiol.* **180**: 155-160.

Elbahloul, Y. & A. Steinbuchel, (2006) Engineering the genotype of *Acinetobacter* sp. strain ADP1 to enhance biosynthesis of cyanophycin. *Appl. Environ. Microbiol.* **72**: 1410-1419.

Emsley, P. & K. Cowtan, (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132.

Ezezika, O. C., L. S. Collier-Hyams, H. A. Dale, A. C. Burk & E. L. Neidle, (2006) CatM regulation of the *benABCDE* operon: functional divergence of two LysR-type paralogs in *Acinetobacter baylyi* ADP1. *Appl. Environ. Microbiol.* **72**: 1749-1758.

Ezezika, O. C., S. Haddad, T. J. Clark, E. L. Neidle & C. Momany, (2007a) Distinct effector-binding sites enable synergistic transcriptional activation by BenM, a LysR-type regulator. *J Mol Biol* **367**: 616-629.

Ezezika, O. C., S. Haddad, E. L. Neidle & C. Momany, (2007b) Oligomerization of BenM, a LysR-type transcriptional regulator: structural basis for the aggregation of proteins in this family. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **63**: 361-368.

Fournier, P. E., D. Vallenet, V. Barbe, S. Audic, H. Ogata, L. Poirel, H. Richet, C. Robert, S. Mangenot, C. Abergel, P. Nordmann, J. Weissenbach, D. Raoult & J. M. Claverie, (2006) Comparative Genomics of Multidrug Resistance in *Acinetobacter baumannii*. *PLoS Genet* **2**: e7.

Francia, M. V. & D. B. Clewell, (2002) Amplification of the tetracycline resistance determinant of pAMalpha1 in *Enterococcus faecalis* requires a site-specific recombination event involving relaxase. *J. Bacteriol.* **184**: 5187-5193.

Francino, M. P., (2005) An adaptive radiation model for the origin of new gene functions. *Nat. Genet.* **37**: 573-577.

Fritsch, P. S., M. L. Urbanowski & 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**: 5539-5550.

Fujii, T., M. Takeo & Y. Maeda, (1997) Plasmid-encoded genes specifying aniline oxidation from *Acinetobacter* sp. strain YAA. *Microbiology* **143 (Pt 1)**: 93-99.

Gaines, G. L., 3rd, L. Smith & E. L. Neidle, (1996) Novel nuclear magnetic resonance spectroscopy methods demonstrate preferential carbon source utilization by *Acinetobacter calcoaceticus*. *J. Bacteriol.* **178**: 6833-6841.

Gallegos, M. T., C. Michan & J. L. Ramos, (1993) The XylS/AraC family of regulators. *Nucleic Acids Res* **21**: 807-810.

Galvao, T. C. and V. de Lorenzo, (2006) Transcriptional regulators a la carte: engineering new effector specificities in bacterial regulatory proteins. *Curr Opin in Biotechnol* **17:** 34-42.

Geissdorfer, W., R. G. Kok, A. Ratajczak, K. J. Hellingwerf & W. Hillen, (1999) The genes *rubA* and *rubB* for alkane degradation in *Acinetobacter* sp. strain ADP1 are in an operon with *estB*, encoding an esterase, and *oxyR*. *J Bacteriol* **181**: 4292-4298.

Gerischer, U., (2002) Specific and global regulation of genes associated with the degradation of aromatic compounds in bacteria. *J Mol Microbiol Biotechnol* **4**: 111-121.

Gevers, D., K. Vandepoele, C. Simillon & Y. Van de Peer, (2004) Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends Microbiol.* **12**: 148-154.

Gey van Pittius, N. C., S. L. Sampson, H. Lee, Y. Kim, P. D. van Helden & R. M. Warren, (2006) Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (*esx*) gene cluster regions. *BMC Evol. Biol.* **6**: 95.

Gourse, R. L., W. Ross & T. Gaal, (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol. Microbiol.* **37**: 687-695.

Gralton, E. M., A. L. Campbell & E. L. Neidle, (1997) Directed introduction of DNA cleavage sites to produce a high-resolution genetic and physical map of the Acinetobacter sp. strain ADP1 (BD413UE) chromosome. *Microbiology* **143 (Pt 4)**: 1345-1357.

Grant, B. J., A. P. Rodrigues, K. M. ElSawy, J. A. McCammon & L. S. Caves, (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**: 2695-2696.

Haack, K. R. & J. R. Roth, (1995) Recombination between chromosomal IS200 elements supports frequent duplication formation in Salmonella typhimurium. *Genetics* **141**: 1245-1252.

Harms, H., M. C. Wells & J. R. van der Meer, (2006) Whole-cell living biosensors--are they ready for environmental application? *Appl Microbiol Biotechnol* **70**: 273-280.

Hastings, P. J., A. Slack, J. F. Petrosino & S. M. Rosenberg, (2004) Adaptive amplification and point mutation are independent mechanisms: evidence for various stress-inducible mutation mechanisms. *PLoS Biol.* **2**: e399.

Hayward, S., (1999) Structural principles governing domain motions in proteins. *Proteins* **36**: 425-435.

Hayward, S. & H. J. Berendsen, (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30**: 144-154.

Hayward, S. & R. A. Lee, (2002) Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model* **21**: 181-183.

Heil, G., L. T. Stauffer & G. V. Stauffer, (2002) Glycine binds the transcriptional accessory protein GcvR to disrupt a GcvA/GcvR interaction and allow GcvA-mediated activation of the *Escherichia coli gcvTHP* operon. *Microbiology* **148**: 2203-2214.

Henikoff, S., G. W. Haughn, J. M. Calvo & J. C. Wallace, (1988) A large family of bacterial activator proteins. *Proc. Natl. Acad. Sci. U.S.A.* **85**: 6602-6606.

Hooper, S. D. & O. G. Berg, (2003) On the nature of gene innovation: duplication patterns in microbial genomes. *Mol. Biol. Evol.* **20**: 945-954.

Hryniewicz, M. M. & N. M. Kredich, (1994) Stoichiometry of binding of CysB to the *cysJIH*, *cysK*, and *cysP* promoter regions of *Salmonella typhimurium*. *J Bacteriol* **176**: 3673-3682.

Hryniewicz, M. M. & N. M. Kredich, (1995) Hydroxyl radical footprints and half-site arrangements of binding sites for the CysB transcriptional activator of *Salmonella typhimurium*. *J Bacteriol* **177**: 2343-2353.

Huang, W. E., L. Huang, G. M. Preston, M. Naylor, J. P. Carr, Y. Li, A. C. Singer, A. S. Whiteley & H. Wang, (2006) Quantitative *in situ* assay of salicylic acid in tobacco leaves using a genetically modified biosensor strain of *Acinetobacter* sp. ADP1. *Plant J.* **46**: 1073-1083.

Huang, W. E., H. Wang, H. Zheng, L. Huang, A. C. Singer, I. Thompson & A. S. Whiteley, (2005) Chromosomally located gene fusions constructed in *Acinetobacter* sp. ADP1 for the detection of salicylate. *Environ. Microbiol.* **7**: 1339-1348.

Huffman, J. L. & R. G. Brennan, (2002) Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr Opin Struct Biol* **12**: 98-106.

Ikeda, H., K. Shiraishi & Y. Ogata, (2004) Illegitimate recombination mediated by double-strand break and end-joining in *Escherichia coli*. *Adv. Biophys.* **38**: 3-20.

Iwanicka-Nowicka, R. & M. M. Hryniewicz, (1995) A new gene, *cbl*, encoding a member of the LysR family of transcriptional regulators belongs to *Escherichia coli cys* regulon. *Gene* **166**: 11-17.

Jomantiene, R. & R. E. Davis, (2006) Clusters of diverse genes existing as multiple, sequence-variable mosaics in a phytoplasma genome. *Fems Microbiol. Lett.* **255**: 59-65.

Jones, R. M., L. S. Collier, E. L. Neidle & P. A. Williams, (1999) *areABC* genes determine the catabolism of aryl esters in *Acinetobacter* sp. Strain ADP1. *J. Bacteriol.* **181**: 4568-4575.

Jones, R. M., V. Pagmantidis & P. A. Williams, (2000) *sal* genes determining the catabolism of salicylate esters are part of a supraoperonic cluster of catabolic genes in *Acinetobacter* sp. strain ADP1. *J. Bacteriol.* **182**: 2018-2025.

Jones, R. M. & P. A. Williams, (2003) Mutational analysis of the critical bases involved in activation of the AreR-regulated sigma54-dependent promoter in *Acinetobacter* sp. strain ADP1. *Appl Environ Microbiol* **69**: 5627-5635.

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

Jovanovic, M., M. Lilic, D. J. Savic & G. Jovanovic, (2003) The LysR-type transcriptional regulator CysB controls the repression of *hslJ* transcription in *Escherichia coli*. *Microbiology* **149**: 3449-3459.

Kahl, B. C., A. Mellmann, S. Deiwick, G. Peters & D. Harmsen, (2005) Variation of the polymorphic region X of the protein A gene during persistent airway infection of cystic fibrosis patients reflects two independent mechanisms of genetic change in *Staphylococcus aureus*. *J. Clin. Microbiol.* **43**: 502-505.

Kapogiannis, B. G., S. Satola, H. L. Keyserling & M. M. Farley, (2005) Invasive infections with *Haemophilus influenzae* serotype a containing an IS1016-*bexA* partial deletion: possible association with virulence. *Clin. Infect. Dis.* **41**: e97-103.

Kim, S. I., Y. C. Yoo & H. Y. Kahng, (2001) Complete nucleotide sequence and overexpression of *cat1* gene cluster, and roles of the putative transcriptional activator CatR1 in *Acinetobacter lwoffii* K24 capable of aniline degradation. *Biochem. Biophys. Res. Commun.* **288**: 645-649.

Kroll, J. S., B. M. Loynds & E. R. Moxon, (1991) The Haemophilus influenzae capsulation gene cluster: a compound transposon. *Mol Microbiol* **5**: 1549-1560.

Kugelberg, E., E. Kofoid, A. B. Reams, D. I. Andersson & J. R. Roth, (2006) Multiple pathways of selected gene amplification during adaptive mutation. *Proc. Natl. Acad. Sci. U.S.A.* **103**: 17319-17324.

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

Larsson, P., P. C. Oyston, P. Chain, M. C. Chu, M. Duffield, H. H. Fuxelius, E. Garcia, G. Halltorp, D. Johansson, K. E. Isherwood, P. D. Karp, E. Larsson, Y. Liu, S. Michell, J. Prior, R. Prior, S. Malfatti, A. Sjostedt, K. Svensson, N. Thompson, L. Vergez, J. K. Wagg, B. W. Wren, L. E. Lindler, S. G. Andersson, M. Forsman & R. W. Titball, (2005) The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. *Nat. Genet.* **37**: 153-159.

Lee, Y., S. Pena-Llopis, Y. S. Kang, H. D. Shin, B. Demple, E. L. Madsen, C. O. Jeon & W. Park, (2006) Expression analysis of the *fpr* (ferredoxin-NADP+ reductase) gene in *Pseudomonas putida KT2440*. *Biochem Biophys Res Commun* **339**: 1246-1254.

Lee, Y. S., H. Kim & D. S. Hwang, (1996) Transcriptional activation of the *dnaA* gene encoding the initiator for *oriC* replication by IciA protein, an inhibitor of in vitro *oriC* replication in *Escherichia coli*. *Mol. Microbiol.* **19**: 389-396.

Lenco, J., M. Hubalek, P. Larsson, A. Fucikova, M. Brychta, A. Macela & J. Stulik, (2007) Proteomics analysis of the *Francisella tularensis* LVS response to iron restriction: induction of the *F. tularensis* pathogenicity island proteins IglABC. *FEMS Microbiol. Lett.*

Lindstedt, B. A., (2005) Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* **26**: 2567-2582.

Lochowska, A., R. Iwanicka-Nowicka, D. Plochocka & M. M. Hryniewicz, (2001) Functional dissection of the LysR-type CysB transcriptional regulator. Regions important for DNA binding, inducer response, oligomerization, and positive control. *J Biol Chem* **276**: 2098-2107.

Lochowska, A., R. Iwanicka-Nowicka, J. Zaim, M. Witkowska-Zimny, K. Bolewska & M. M. Hryniewicz, (2004) Identification of activating region (AR) of *Escherichia coli* LysR-type transcription factor CysB and CysB contact site on RNA polymerase alpha subunit at the *cysP* promoter. *Mol Microbiol* **53**: 791-806.

Lonneborg, R., I. Smirnova, C. Dian, G. A. Leonard & P. Brzezinski, (2007) *In vivo* and *in vitro* investigation of transcriptional regulation by DntR. *J Mol Biol* **372**: 571-582.

Lovett, S. T., (2004) Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* **52**: 1243-1253.

Luft, J. R., J. Wofley, I. Jusrisica, J. Glasgow, S. Fortier & G. T. DeTitta, (2001) Macromolecular crystallization in a high throughput laboratory- the search phase. *J Crystal Growth* **232**: 591-595.

Ma, B., M. Shatsky, H. J. Wolfson & R. Nussinov, (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci* **11**: 184-197.

Martinez-Bueno, M., A. J. Molina-Henares, E. Pareja, J. L. Ramos & R. Tobes, (2004) BacTregulators: a database of transcriptional regulators in bacteria and archaea. *Bioinformatics* **20**: 2787-2791.

McFall, S. M., S. A. Chugani & A. M. Chakrabarty, (1998) Transcriptional activation of the catechol and chlorocatechol operons: variations on a theme. *Gene* **223**: 257-267.

Miller, J. H., (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Monroe, R. S., J. Ostrowski, M. M. Hryniewicz & N. M. Kredich, (1990) In vitro interactions of CysB protein with the *cysK* and *cysJIH* promoter regions of *Salmonella typhimurium*. *J Bacteriol* **172**: 6919-6929.

Mrazek, J., (2006) Analysis of distribution indicates diverse functions of simple sequence repeats in *Mycoplasma* genomes. *Mol. Biol. Evol.* **23**: 1370-1385.

Mrazek, J. & S. Xie, (submitted) Pattern Locator: A tool for finding local sequence patterns in genomic DNA sequences. *Bioinformatics*.

Mulder, N. J., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bradley, P. Bork, P. Bucher, L. Cerutti, R. Copley, E. Courcelle, U. Das, R. Durbin, W. Fleischmann,

J. Gough, D. Haft, N. Harte, N. Hulo, D. Kahn, A. Kanapin, M. Krestyaninova, D. Lonsdale, R. Lopez, I. Letunic, M. Madera, J. Maslen, J. McDowall, A. Mitchell, A. N. Nikolskaya, S. Orchard, M. Pagni, C. P. Ponting, E. Quevillon, J. Selengut, C. J. Sigrist, V. Silventoinen, D. J. Studholme, R. Vaughan & C. H. Wu, (2005) InterPro, progress and status in 2005. *Nucleic Acids Res.* **33**: D201-205.

Murakami, S., T. Hayashi, T. Maeda, S. Takenaka & K. Aoki, (2003) Cloning and functional analysis of aniline dioxygenase gene cluster, from *Frateuria* species ANA-18, that metabolizes aniline via an *ortho*-cleavage pathway of catechol. *Biosci. Biotechnol. Biochem.* **67**: 2351-2358.

Muraoka, S., R. Okumura, N. Ogawa, T. Nonaka, K. Miyashita & T. Senda, (2003) Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *J Mol Biol* **328**: 555-566.

Murshudov, G. N., A. A. Vagin & E. J. Dodson, (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255.

Musher, D. M., M. E. Dowell, V. D. Shortridge, R. K. Flamm, J. H. Jorgensen, P. Le Magueres & K. L. Krause, (2002) Emergence of macrolide resistance during treatment of pneumococcal pneumonia. *N. Engl. J. Med.* **346**: 630-631.

Mussi, M. A., A. S. Limansky & A. M. Viale, (2005) Acquisition of resistance to carbapenems in multidrug-resistant clinical strains of *Acinetobacter baumannii*: natural insertional inactivation of a gene encoding a member of a novel family of beta-barrel outer membrane proteins. *Antimicrob. Agents Chemother.* **49**: 1432-1440.

Nair, S., D. Nash, D. Sudimack, M. Barends, A-C Uhlemann, S. Krishna, F. Nosten, T. J. C. Anderson. (2006). Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol. Biol. Evol.* **24**: 562-573.

Nakada, Y. & Y. Itoh, (2002) Characterization and regulation of the *gbuA* gene, encoding guanidinobutyrase in the arginine dehydrogenase pathway of *Pseudomonas aeruginosa* PAO1. *J Bacteriol* **184**: 3377-3384.

Nandineni, M. R. & J. Gowrishankar, (2004) Evidence for an arginine exporter encoded by *yggA* (*argO*) that is regulated by the LysR-type transcriptional regulator ArgP in *Escherichia coli*. *J. Bacteriol.* **186**: 3539-3546.

Neidle, E. L., C. Hartnett & L. N. Ornston, (1989) Characterization of *Acinetobacter calcoaceticus catM*, a repressor gene homologous in sequence to transcriptional activator genes. *J Bacteriol* **171**: 5410-5421.

Neidle, E. L., M. K. Shapiro & L. N. Ornston, (1987) Cloning and expression in *Escherichia coli* of *Acinetobacter calcoaceticus* genes for benzoate degradation. *J. Bacteriol.* **169**: 5496-5503.

Nicoloff, H., V. Perreten & S. B. Levy, (2007) Increased genome instability in Escherichia coli lon mutants: relation to emergence of multiple-antibiotic-resistant (Mar) mutants caused by insertion sequence elements and large tandem genomic amplifications. *Antimicrob Agents Chemother* **51**: 1293-1303.

Nicoloff, H., V. Perreten, L. M. McMurry & S. B. Levy, (2006) Role for tandem duplication and lon protease in AcrAB-TolC- dependent multiple antibiotic resistance (Mar) in an *Escherichia coli* mutant without mutations in *marRAB* or *acrRAB*. *J. Bacteriol.* **188**: 4413-4423.

Nilsson, A. I., A. Zorzet, A. Kanth, S. Dahlstrom, O. G. Berg & D. I. Andersson, (2006) Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. *Proc. Natl. Acad. Sci. U.S.A.* **103**: 6976-6981.

Otwinowski, Z. & W. Minor, (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307-326.

Papazisi, L., T. S. Gorton, G. Kutish, P. F. Markham, G. F. Browning, D. K. Nguyen, S. Swartzell, A. Madan, G. Mahairas & S. J. Geary, (2003) The complete genome sequence of the avian pathogen *Mycoplasma gallisepticum* strain R(low). *Microbiology* **149**: 2307-2316.

Pareja, E., P. Pareja-Tobes, M. Manrique, E. Pareja-Tobes, J. Bonal & R. Tobes, (2006) ExtraTrain: a database of Extragenic regions and Transcriptional information in prokaryotic organisms. *BMC Microbiol* **6**: 29.

Parke, D., (1996) Conservation of PcaQ, a transcriptional activator of pca genes for catabolism of phenolic compounds, in *Agrobacterium tumefaciens* and *Rhizobium* species. *J. Bacteriol.* **178**: 3671-3675.

Parke, D. & L. N. Ornston, (2003) Hydroxycinnamate (*hca*) catabolic genes from *Acinetobacter* sp. strain ADP1 are repressed by HcaR and are induced by hydroxycinnamoyl-coenzyme A thioesters. *Appl. Environ. Microbiol.* **69**: 5398-5409.

Parsek, M. R., M. Kivisaar & A. M. Chakrabarty, (1995) Differential DNA bending introduced by the *Pseudomonas putida* LysR-type regulator, CatR, at the plasmid-borne *pheBA* and chromosomal *catBC* promoters. *Mol Microbiol* **15**: 819-828.

Peck, M. C., R. F. Fisher & S. R. Long, (2006) Diverse Flavonoids Stimulate NodD1 Binding to *nod* Gene Promoters in *Sinorhizobium meliloti*. *J Bacteriol* **188**: 5417-5427.

Peng, H. L., S. R. Shiou & H. Y. Chang, (1999) Characterization of *mdcR*, a regulatory gene of the malonate catabolic system in *Klebsiella pneumoniae*. *J. Bacteriol.* **181**: 2302-2306.

Pitcher, R. S., N. C. Brissett & A. J. Doherty, (2007) Nonhomologous end-joining in bacteria: a microbial perspective. *Annu Rev Microbiol* **61**: 259-282.

Ponder, R. G., N. C. Fonville & S. M. Rosenberg, (2005) A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Mol. Cell.* **19**: 791-804.

Quiocho, F. A. & P. S. Ledvina, (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* **20**: 17-25.

Raschke, H., M. Meier, J. G. Burken, R. Hany, M. D. Muller, J. R. Van Der Meer & H. P. Kohler, (2001) Biotransformation of various substituted aromatic compounds to chiral dihydrodihydroxy derivatives. *Appl. Environ. Microbiol.* **67**: 3333-3339.

Reams, A. B. & E. L. Neidle, (2003) Genome plasticity in Acinetobacter: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. *Mol Microbiol* **47**: 1291-1304.

Reams, A. B. & E. L. Neidle, (2004a) Gene amplification involves site-specific short homology-independent illegitimate recombination in *Acinetobacter* sp. strain ADP1. *J. Mol. Biol.* **338**: 643-656.

Reams, A. B. & E. L. Neidle, (2004b) Selection for gene clustering by tandem duplication. *Annu. Rev. Microbiol.* **58**: 119-142.

Rocha, E. P. & A. Blanchard, (2002) Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution. *Nucleic Acids Res.* **30**: 2031-2042.

Romero, D. & R. Palacios, (1997) Gene amplification and genomic plasticity in prokaryotes. *Annu. Rev. Genet.* **31**: 91-111.

Romero-Arroyo, C. E., M. A. Schell, G. L. Gaines, 3rd & E. L. Neidle, (1995) *catM* encodes a LysR-type transcriptional activator regulating catechol degradation in *Acinetobacter calcoaceticus*. *J. Bacteriol.* **177**: 5891-5898.

Rosario, C. J. & R. A. Bender, (2005) Importance of tetramer formation by the nitrogen assimilation control protein for strong repression of glutamate dehydrogenase formation in *Klebsiella pneumoniae*. *J Bacteriol* **187**: 8291-8299.

Roth, J. R., E. Kugelberg, A. B. Reams, E. Kofoid & D. I. Andersson, (2006) Origin of mutations under selection: the adaptive mutation controversy. *Annu. Rev. Microbiol.* **60**: 477-501.

Sambrook, J., E. F. Fritsch & T. Maniatis, (1989) *Molecular cloning: a laboratory manual, 2nd ed.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.

Sainsbury, S., L. A. Lane, C. V. Robinson, D. I. Stuart, & R. J. Owens, (2009) The structure of CrgA from *Neisseria meningitidis* reveals a new octameric assembly state for LysR transcriptional regulators. *Nucleic Acids Res.* **in press**.

Sayle, R. A. & E. J. Milner-White, (1995) RASMOL: biomolecular graphics for all. *Trends Biochem Sci* **20**: 374.

Schell, M. A., (1993) Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol.* **47**: 597-626.

Schell, M. A., P. H. Brown & S. Raju, (1990) Use of saturation mutagenesis to localize probable functional domains in the NahR protein, a LysR-type transcription activator. *J Biol Chem* **265**: 3844-3850.

Seoane, A., E. Sanchez & J. M. Garcia-Lobo, (2003) Tandem amplification of a 28-kilobase region from the *Yersinia enterocolitica* chromosome containing the *blaA* gene. *Antimicrob. Agents Chemother.* **47**: 682-688.

Seshadri, R., L. Adrian, D. E. Fouts, J. A. Eisen, A. M. Phillippy, B. A. Methe, N. L. Ward, W. C. Nelson, R. T. Deboy, H. M. Khouri, J. F. Kolonay, R. J. Dodson, S. C. Daugherty, L. M. Brinkac, S. A. Sullivan, R. Madupu, K. E. Nelson, K. H. Kang, M. Impraim, K. Tran, J. M. Robinson, H. A. Forberger, C. M. Fraser, S. H. Zinder & J. F. Heidelberg, (2005) Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. *Science* **307**: 105-108.

Shanley, M. S., E. L. Neidle, R. E. Parales & L. N. Ornston, (1986) Cloning and expression of Acinetobacter calcoaceticus catBCDE genes in Pseudomonas putida and Escherichia coli. *J Bacteriol* **165**: 557-563.

Siehler, S. Y., S. Dal, R. Fischer, P. Patz & U. Gerischer, (2007) Multiple-level regulation of genes for protocatechuate degradation in Acinetobacter baylyi includes cross-regulation. *Appl Environ Microbiol* **73**: 232-242.

Siguier, P., J. Filee & M. Chandler, (2006) Insertion sequences in prokaryotic genomes. *Curr. Opin. Microbiol.* **9**: 526-531.

Simmons, W. L., J. R. Bolland, J. M. Daubenspeck & K. Dybvig, (2006) A stochastic mechanism for biofilm formation by *Mycoplasma pulmonis*. *J. Bacteriol.* ***189****:* 1905-1913.

Simmons, W. L., A. M. Denison & K. Dybvig, (2004) Resistance of *Mycoplasma pulmonis* to complement lysis is dependent on the number of Vsa tandem repeats: shield hypothesis. *Infect. Immun.* **72**: 6846-6851.

Simmons, W. L. & K. Dybvig, (2003) The Vsa proteins modulate susceptibility of *Mycoplasma pulmonis* to complement killing, hemadsorption, and adherence to polystyrene. *Infect. Immun.* **71**: 5733-5738.

Singer, A. C., D. E. Crowley & I. P. Thompson, (2003) Secondary plant metabolites in phytoremediation and biotransformation. *Trends Biotechnol.* **21**: 123-130.

Slack, A., P. C. Thornton, D. B. Magner, S. M. Rosenberg & P. J. Hastings, (2006) On the mechanism of gene amplification induced under stress in *Escherichia coli*. *PLoS Genet.* **2**: e48.

Smart, J. L. & C. E. Bauer, (2006) Tetrapyrrole biosynthesis in *Rhodobacter capsulatus* is transcriptionally regulated by the heme-binding regulatory protein, HbrL. *J Bacteriol* **188**: 1567-1576.

Smirnova, I. A., C. Dian, G. A. Leonard, S. McSweeney, D. Birse & P. Brzezinski, (2004) Development of a bacterial biosensor for nitrotoluenes: the crystal structure of the transcriptional regulator DntR. *J Mol Biol* **340**: 405-418.

Stauffer, L. T. & G. V. Stauffer, (2005) GcvA interacts with both the alpha and sigma subunits of RNA polymerase to activate the *Escherichia coli gcvB* gene and the *gcvTHP* operon. *FEMS Microbiol. Lett.* **242**: 333-338.

Stec, E., M. Witkowska-Zimny, M. M. Hryniewicz, P. Neumann, A. J. Wilkinson, A. M. Brzozowski, C. S. Verma, J. Zaim, S. Wysocki & G. D. Bujacz, (2006) Structural basis of the sulphate starvation response in *E. coli*: crystal structure and mutational analysis of the cofactor-binding domain of the Cbl transcriptional regulator. *J Mol Biol* **364**: 309-322.

Stumpf, J. D., A. R. Poteete & P. L. Foster, (2007) Amplification of *lac* cannot account for adaptive mutation to Lac+ in *Escherichia coli*. *J. Bacteriol.* **189**: 2291-2299.

Sun, S., O. G. Berg, J. R. Roth, and D. I. Anderson, (2009) Contribution of gene amplification to evolution of increased antibiotic resistance in *Salmonella typhimurium*. *Genetics*. **in press.**

Swanson, J. A., J. T. Mulligan & S. R. Long, (1993) Regulation of *syrM* and *nodD3* in *Rhizobium meliloti*. *Genetics* **134**: 435-444.

Tao, K., C. Zou, N. Fujita & A. Ishihama, (1995) Mapping of the OxyR protein contact site in the C-terminal region of RNA polymerase alpha subunit. *J. Bacteriol.* **177**: 6740-6744.

Tlsty, T. D., A. M. Albertini & J. H. Miller, (1984) Gene amplification in the *lac* region of *E. coli*. *Cell* **37**: 217-224.

Tropel, D. & J. R. van der Meer, (2004) Bacterial transcriptional regulators for degradation pathways of aromatic compounds. *Microbiol Mol Biol Rev* **68**: 474-500.

Tyrrell, R., K. H. Verschueren, E. J. Dodson, G. N. Murshudov, C. Addy & A. J. Wilkinson, (1997) The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Structure* **5**: 1017-1032.

Urata, M., E. Uchida, H. Nojiri, T. Omori, R. Obo, N. Miyaura & N. Ouchiyama, (2004) Genes involved in aniline degradation by Delftia acidovorans strain 7N and its distribution in the natural environment. *Biosci Biotechnol Biochem* **68**: 2457-2465.

Valderramos, S. G. & D. A. Fidock, (2006) Transporters involved in resistance to antimalarial drugs. *Trends Pharmacol. Sci.* **27**: 594-601.

van Belkum, A., (2007) Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). *FEMS Immunol. Med. Microbiol.* **49**: 22-27.

van der Meer, J. R., D. Tropel & M. Jaspers, (2004) Illuminating the detection chain of bacterial bioreporters. *Environ Microbiol* **6**: 1005-1020.

van der Ploeg, J. R., E. Eichhorn & T. Leisinger, (2001) Sulfonate-sulfur metabolism and its regulation in *Escherichia coli*. *Arch. Microbiol.* **176**: 1-8.

Vaneechoutte, M., D. M. Young, L. N. Ornston, T. De Baere, A. Nemec, T. Van Der Reijden, E. Carr, I. Tjernberg & L. Dijkshoorn, (2006) Naturally transformable *Acinetobacter* sp. strain ADP1 belongs to the newly described species *Acinetobacter baylyi*. *Appl Environ Microbiol* **72**: 932-936.

Verschueren, K. H., R. Tyrrell, G. N. Murshudov, E. J. Dodson & A. J. Wilkinson, (1999) Solution of the structure of the cofactor-binding fragment of CysB: a struggle against non-isomorphism. *Acta Crystallogr D Biol Crystallogr* **55**: 369-378.

Vogler, A. J., C. Keys, Y. Nemoto, R. E. Colman, Z. Jay & P. Keim, (2006) Effect of repeat copy number on variable-number tandem repeat mutations in *Escherichia coli* O157:H7. *J. Bacteriol.* **188**: 4253-4263.

Vogler, A. J., C. E. Keys, C. Allender, I. Bailey, J. Girard, T. Pearson, K. L. Smith, D. M. Wagner & P. Keim, (2007) Mutations, mutation rates, and evolution at the hypervariable VNTR loci of *Yersinia pestis*. *Mutat. Res.* **616**: 145-158.

Waite, R. D., D. W. Penfold, J. K. Struthers & C. G. Dowson, (2003) Spontaneous sequence duplications within capsule genes *cap8E* and *tts* control phase variation in *Streptococcus pneumoniae* serotypes 8 and 37. *Microbiology* **149**: 497-504.

Waite, R. D., J. K. Struthers & C. G. Dowson, (2001) Spontaneous sequence duplication within an open reading frame of the pneumococcal type 3 capsule locus causes high-frequency phase variation. *Mol. Microbiol.* **42**: 1223-1232.

Wang, L., J. D. Helmann & S. C. Winans, (1992) The *A. tumefaciens* transcriptional activator OccR causes a bend at a target promoter, which is partially relaxed by a plant tumor metabolite. *Cell* **69**: 659-667.

Williams, P. A. & L. E. Shaw, (1997) *mucK*, a gene in *Acinetobacter calcoaceticus* ADP1 (BD413), encodes the ability to grow on exogenous *cis,cis*-muconate as the sole carbon source. *J. Bacteriol.* **179**: 5935-5942.

Xiao, G., J. He & L. G. Rahme, (2006) Mutation analysis of the *Pseudomonas aeruginosa mvfR* and *pqsABCDE* gene promoters demonstrates complex quorum-sensing circuitry. *Microbiology* **152**: 1679-1686.

Yanai, K., T. Murakami & M. Bibb, (2006) Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of *Streptomyces kanamyceticus*. *Proc. Natl. Acad. Sci. U.S.A* **103**: 9661-9666.

Yang, S. J., P. M. Dunman, S. J. Projan & K. W. Bayles, (2006) Characterization of the *Staphylococcus aureus* CidR regulon: elucidation of a novel role for acetoin metabolism in cell death and lysis. *Mol Microbiol* **60**: 458-468.

Yasui, K., S. Mihara, C. Zhao, H. Okamoto, F. Saito-Ohara, A. Tomida, T. Funato, A. Yokomizo, S. Naito, I. Imoto, T. Tsuruo & J. Inazawa, (2004) Alteration in copy numbers of genes as a mechanism for acquired drug resistance. *Cancer Res.* **64**: 1403-1410.

Young, D. M., D. Parke & L. N. Ornston, (2005) Opportunities for genetic investigation afforded by *Acinetobacter baylyi*, a nutritionally versatile bacterial species that is highly competent for natural transformation. *Annu Rev Microbiol* **59**: 519-551.

Zhan, Y., H. Yu, Y. Yan, M. Chen, W. Lu, S. Li, Z. Peng, W. Zhang, S. Ping, J. Wang & M. Lin, (2008) Genes Involved in the Benzoate Catabolic Pathway in *Acinetobacter calcoaceticus* PHEA-2. *Curr Microbiol* **57**: 609-614.

Zhang, J., (2003) Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**: 292-298.

APPENDIX A

SUPPLEMENTARY STRUCTURAL DATA FOR BENM AND CATM VARIANT STRUCTURES

Table A1: Calculated values describing the domain motions for BenM-EBD structures.

Structure and chain versus fixed BenM structure 2F6G monomer A	2F6G-B	2F7A-A	2F7A-B	2H9B-A	2H9B-B	2H99-A	2H99-B	2F8D-A	2F8D-B	2F97-A	2F78-A	2F78-B
Note	wild type low pH1	wild type low pH1	wild type low pH1	R156, T157S variant	R156, T157S variant	R156, T157S variant	R156,T 157S variant	wild type high pH1	wild type high pH1	wild type high pH1	wild type low pH1	wild type low pH1
Primary Site Ligand	none	none	muconate	1 SO4	1 SO4 2, 1 Cl	1 acetate	1 SO4 - 2, 1 Cl	benzoate	none	none	benzoate	benzoate
Secondary Site Ligand	none	benzoate	benzoate	1 Cl	none	1 acetate	1 acetate	benzoate	benzoate	acetate	benzoate	benzoate
RMSD Of Fixed Domain I Best Fit (Å)	0.663	0.657	1.031	1.102	0.879	1.050	1.248	1.053	1.019	1.064	0.842	1.18
Percentage External Displacement In Moving Domain (%)	94.761	93.224	68.467	92.553	91.486	93.484	91.885	94.116	75.000	74.461	96.737	97.327
Angle of Rotation (°)	3.989	6.651	5.481	2.695	4.336	2.709	4.022	3.711	3.792	4.267	4.996	4.163
Translation Along Axis (Å)	-0.209	-0.644	0.209	0.700	0.359	0.649	0.366	0.657	0.547	0.316	-0.070	0.224
Angle Between Screw Axis And Line Joining Centres Of Mass (°)	37.159	46.242	83.920	67.678	39.635	75.095	40.260	96.606	121.362	110.555	61.461	110.652
Distance Between Screw Axis And Line Joining Centres Of Mass (Å)	0.309	4.832	10.461	16.44	19.87	14.371	22.269	3.436	1.150	4.493	1.488	5.789
Percentage Closure Motion (%)	36.486	52.166	98.878	85.574	40.691	93.384	41.764	98.676	72.914	87.673	77.175	87.561
Normalized Screw Axis Vector[1]	0.229, -0.873, -0.431	-0.127, -0.741, -0.659	-0.544, 0.034, 0.839	-0.180, -0.153, 0.972	-0.105, -0.697, 0.709	-0.092, -0.289, 0.953	-0.119, -0.703, 0.701	0.433, -0.016, -0.901	0.657, 0.494, 0.569	0.418, 0.375, 0.827	-0.008, -0.555, -0.832	-0.230, 0.305, -0.924
Screw Axis Point2	14.221, 23.089, 42.664	12.298, 26.495, 39.234	8.670, 37.742, 34.174	1.651, 37.208, 42.710	-2.489, 33.482, 48.950	3.882, 40.142, 39.964	-4.606, 34.169, 49.313	14.839, 19.037, 34.764	12.642, 37.707, 33.727	12.835, 38.849, 35.903	15.373, 24.512, 39.999	10.878, 26.471, 38.523
RMSD of Rotated Domain II at Fixed position (Å)	0.974	1.757	1.761	1.230	1.650	1.130	1.689	1.763	0.990	0.999	1.451	1.166

[1]The normalized screw axis vector and axis points were derived from the DynDom (Hayward & Berendsen, 1998, Hayward & Lee, 2002) output used to model the vectors in the program Rasmol (Sayle & Milner-White, 1995). A python script was written to calculate these values using the program Pymol (DeLano, 2002). Axes are relative to the A subunit of structure 2F6G.

Table A2: Calculated values describing the domain motions for CatM-EBD structures.

Structure and chain versus fixed CatM structure 2F7B[1]	2F7C-A	3GLB-A	3GLB-B	3GLB-C	3GLB-D
Note	wild type	R156H	R156H	R156H	R156H
Primary Site Ligand	muconate	muconate	muconate	muconate	muconate
RMSD of Fixed Domain I Best Fit (Å)	0.394	0.640	0.628	0.670	0.649
Percentage External Displacement in Moving Domain (%)	93.220	53.362	53.566	60.702	58.505
Angle of Rotation (°)	4.251	5.646	4.225	7.500	6.885
Translation Along Axis (Å)	-0.418	-0.347	-0.349	-0.541	-0.354
Angle Between Screw Axis And Line Joining Centres of Mass (°)	73.550	33.804	32.623	18.408	26.517
Distance Between Screw Axis And Line Joining Centres of Mass (Å)	2.452	5.458	4.867	0.050	1.272
Percentage Closure Motion (%)	91.981	30.953	29.064	9.971	19.934
Normalized Screw Axis Vector	-0.362, -0.588, -0.723	-0.826, -0.112, -0.553	-0.687, 0.724, 0.055	-0.988, -0.051, -0.147	-0.971, -0.191, -0.144
Screw Axis Point	63.517, 14.313, 63.024	57.041, 12.730, 62.281	57.542, 15.617, 56.139	57.813, 16.861, 62.398	58.589, 17.138, 63.893
RMSD of Rotated Domain II at Fixed position (Å)	1.375	1.755	1.907	1.826	1.702

[1] 2F7B has one subunit per asymmetric unit and contains 1 SO_4^{-2} and 1 Cl^- in the primary site[2] The normalized screw axis vector and axis points were derived from the DynDom (Hayward & Berendsen, 1998, Hayward & Lee, 2002) output used to model the vectors in the program Rasmol (Sayle & Milner-White, 1995). A python script was written to calculate these values using the program Pymol (DeLano, 2002). Axes are relative to structure 2F7B.

Table A.3: Primers to introduce amino acid substitutions at positions 156, 160, 225, 226 or 156

Primer	Sequence [1]
BenMY293F1	5'- CGGCTCAGTAAAACCTTCATAGGC**ga**A**A**TATCTGGCGAATC -3'
BenMY293F2	5'- GATTCGCCAGATA**Ttc**GCCTATGAAGGTTTTACTGAGCCG -3'
BenMR160H1	5'- CGGCTACCATCAAACGTTCATT**AtG**CAATAAGG -3'
BenMR160H2	5'- CCTTATTG**CaT**AATGAACGTTTGATGGTAGCCG -3'
BenMR160K1	5'- CGGCTACCATCAAACGTTCATT**ctt**CAATAAGG -3'
BenMR160K2	5'- CCTTATTG**aag**AATGAACGTTTGATGGTAGCCG -3'
BenMR160M1	5'- CGGCTACCATCAAACGTTCATT**cat**CAATAAGG -3'
BenMR160M2	5'- CCTTATTG**atg**AATGAACGTTTGATGGTAGCCG -3'
BenMR225H1	5'-CAAGCGCCAGTTGAACTTC**AtG**GACTTCATTAATTTTGGTG-3'
BenMR225H2	5'-CACCAAAATTAATGAAGTC**CaT**GAAGTTCAACTGGCGCTTG-3'
BenME226K1	5'-CCAAGCGCCAGTTGAAC**TTt**ACGGACTTCATTAATTTTG-3'
BenME226K2	5'-CAAAATTAATGAAGTCCGT**aAA**GTTCAACTGGCGCTTGG-3'
SC1-BenMR156H	5'- CAGTGATCCCGCCATTAAA**CaT**ACCTTATTGC -3'
SC2-BenMR156H	5'- GCAATAAGGT**AtG**TTTAATGGCGGGATCACTG -3'
R156HB-FR-QC	5'- CAGTGATCCCGCCATTAAA**CaT**tCCTTATTGCGTAATG -3'
R156HB-RV-QC	5'- CATTACGCAATAAGGa**AtG**TTTAATGGCGGGATCACTG -3'

[1]Nucleotides highlighted in bold text encode the desired amino acid change. Nucleotides shown in lowercase text represent bases that were altered and thus do not correspond to the genome sequence (GenBank Accession Number NC005966).

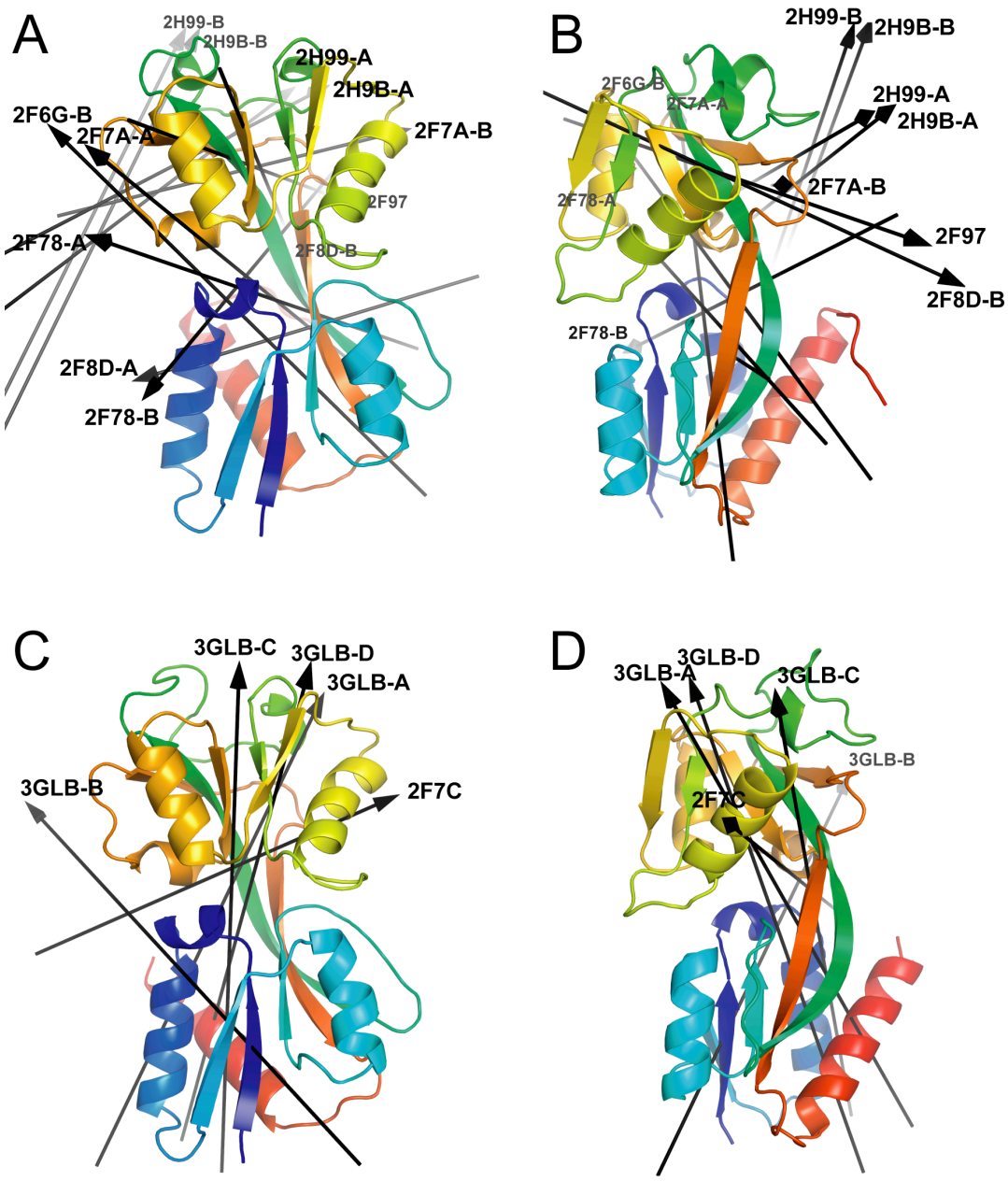


Figure A.1: Screw axes denoting domain motion for available structures of BenM and CatM effector binding domains. Panel A shows the screw axes overlaid on a ribbon representation of the unliganded BenM-EBD structure (PDB Id 2F6G monomer A) colored blue to red, amino terminus to carboxy terminus. Panel B is the same figure (BenM-EBD) as Panel A rotated roughly 90º about the perpendicular axis. Panel C shows a similar representation of the unliganded CatM-EBD subunit structure (PDB Id 2F7B) and the corresponding screw axes of CatM-EBD structures. Panel D is Panel C rotated roughly 90º. The PDB ID and specific subunit corresponding to each structure's screw axis is shown in bold text at the end of each axis. Screw axes were visualized by reformatting the Rasmol-format allowing visualization in the program Pymol (Delano, 2002).